

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 4, 2004, 14:31:59 ; Search time 3149 Seconds
(without alignments)
2962.018 Million cell updates/sec

Title: US-09-675-650-1_COPY_27_254

Perfect score: 228

Sequence: 1 gtaagcgtctcataagcac.....tttcctccctccattccacag 228

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched:

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_ph:*

7: gb_pl:*

8: gb_pr:*

9: gb_ro:*

10: gb_ets:*

11: gb_ey:*

12: gb_un:*

13: gb_vi:*

14: gb_ba:*

15: em_fun:*

16: em_hum:*

17: em_in:*

18: em_mu:*

19: em_or:*

20: em_ov:*

21: em_pat:*

22: em_ph:*

23: em_pl:*

24: em_ro:*

25: em_ets:*

26: em_un:*

27: em_vl:*

28: em_htg_hum:*

29: em_htg_inv:*

30: em_htg_mus:*

31: em_htg_other:*

32: em_htg_mus:*

33: em_htg_pln:*

34: em_htg_rnd:*

35: em_htg_mam:*

36: em_htg_vrt:*

37: em_sy:*

38: em_htgo_hum:*

39: em_htgo_mus:*

40: em_htgo_other:*

41: em_htgo_mus:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	228	100.0	506	6 AX107857	AX107857 Sequence
2	228	100.0	128480	9 AL359314	AL359314 Human DNA
3	160	70.2	5435	9 AF103908	AF103908 Homo sapi
4	23	10.1	224052	2 AC124316	AC124316 Mus muscu
5	22	9.6	198517	2 AC115779	AC115779 Mus muscu
6	21	9.2	175577	5 AL929582	AL929582 Zebrafish
7	21	9.2	177658	5 AL935269	AL935269 Zebrafish
8	20	8.8	20	6 AX107860	AX107860 Sequence
9	20	8.8	153479	2 AC087595	AC087595 Oryza sat
10	20	8.8	162161	8 AP003451	AP003451 Oryza sat
11	20	8.8	208430	10 AC122812	AC122812 Mus muscu
12	20	8.8	215616	2 AC109025	AC109025 Rattus no
13	20	8.8	230279	2 AC132315	AC132315 Mus muscu
14	20	8.8	237234	2 AC141860	AC141860 Gallus ga
15	20	8.8	238817	2 AC134221	AC134221 Rattus no
16	20	8.8	243174	2 AC097028	AC097028 Rattus no
17	20	8.8	247309	10 AC122189	AC122189 Mus muscu
18	19	8.3	1717	9 AB050528	AB050528 Macaca fa
19	19	8.3	10084	1 AE010268	AE010268 Pyrococu
20	19	8.3	13588	2 AC012804	AC012804 Drosophi
21	19	8.3	51896	2 AC068031	AC068031 Homo sapi
22	19	8.3	70946	9 AL161614	AL161614 Homo sapi
23	19	8.3	146952	9 AC068522	AC068522 Homo sapi
24	19	8.3	148385	9 HSJ520818	AL121978 Human DNA
25	19	8.3	154814	9 AC093533	AC093533 Homo sapi
26	19	8.3	151304	2 AC016174	AC016174 Homo sapi
27	19	8.3	162939	5 AL954849	AL954849 Zebrafish
28	19	8.3	164396	9 AC108034	AC108034 Homo sapi
29	19	8.3	176785	9 AC116849	AC116849 Mus muscu
30	19	8.3	186978	9 AP005228	AP005228 Homo sapi
31	19	8.3	192409	2 AC072033	AC072033 Homo sapi
32	19	8.3	201886	9 AC092700	AC092700 Homo sapi
33	19	8.3	202983	2 AC024249	AC024249 Homo sapi
34	19	8.3	208618	2 AC019313	AC019313 Homo sapi
35	19	8.3	216488	2 AC109300	AC109300 Mus muscu
36	19	8.3	222692	2 AC123450	AC123450 Rattus no
37	19	8.3	228638	2 AC127208	AC127208 Rattus no
38	19	8.3	228802	2 AC023717	AC023717 Drosophi
39	19	8.3	229077	2 AC096231	AC096231 Rattus no
40	19	8.3	229140	2 AC095567	AC095567 Rattus no
41	19	8.3	233163	2 AC113904	AC113904 Rattus no
42	19	8.3	235182	2 AC099082	AC099082 Rattus no
43	19	8.3	235700	2 AC126298	AC126298 Rattus no
44	19	8.3	242803	2 AC108557	AC108557 Rattus no
45	19	8.3	245082	2 AC097237	AC097237 Rattus no

ALIGNMENTS

RESULT 1

AX107857

LOCUS AX107857

DEFINITION Sequence 1 from Patent WO0123550.

ACCESSION AX107857

VERSION AX107857.1 GI:13923249

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Busse, U., Chypre, C. and Fradet, Y.

PCa3 messenger rna species in benign and malignant prostate tissues

PATENT: WO 0123550-A 1 05-APR-2001;

JOURNAL

FEATURES	Diagnocure Inc. (CA)
SOURCE	Location/Qualifiers
1..506	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
BASE COUNT	132 a 123 c 108 g 143 t
ORIGIN	
Query Match	100.0%; Score 228; DB 6; Length 506;
Best Local Similarity	100.0%; Pred. No. 1.8e-114;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 GTAAGTCTTTATTAAGCAGCATCTATTTCTACTCAGAAATTTTGAATGGCCTTAAGTTCCT 60
Db	
27 GTAAGTCTTTATTAAGCAGCATCTATTTCTACTCAGAAATTTTGAATGGCCTTAAGTTCCT 86	
Qy	61 CTACTCGTTTCTATCCCTCTCTACGACAGTGCCTCCCGGATCCAGTACCGATTTTCTATT 120
Db	
87 CTACTCGTTTCTATCCCTCTCTACGACAGTGCCTCCCGGATCCAGTACCGATTTTCTATT 146	
Qy	121 TCTTGCCCTCGATTGTGCTGACTGGCTCACTTGATTTATTCCTCAGAGAGTGTGATTTTC 180
Db	
147 TCTTGCCCTCGATTGTGCTGACTGGCTCACTTGATTTATTCCTCAGAGAGTGTGATTTTC 206	
Qy	181 TACCCGGGCTCACTTCGCTCCTCCATATTGTGCTCCACATTTTCACAG 228
Db	
207 TACCCGGGCTCACTTCGCTCCTCCATATTGTGCTCCACATTTTCACAG 254	
RESULT 2	
AL359314	128480 bp DNA linear PRI 15-MAY-2001
LOCUS	Human DNA sequence from clone RP11-108L4 on chromosome 9, complete
DEFINITION	sequence.
ACCESSION	AL359314
VERSION	AL359314.14 GI:14132938
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 128480)
JOURNAL	Leongamornlert,D. Direct Submission Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgvery@sanger.ac.uk On May 16, 2001 this sequence version replaced gi:13897374.
COMMENT	4/301 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SwissProt; Tr., TREMBL; Wp., WORMPEP. Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9 RP11-108L4 is from the library RP11-11.1 constructed by the group of Pileter de Jong. For further details see

```

http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-108L4 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-108L4 is at 128480 in this
sequence. The true left end of clone RP11-580J is at 93909 in this
sequence. The true right end of clone RP11-214N6 is at 100 in this
sequence.
FEATURES
      source
      1..128480
         /organism="Homo sapiens"
         /mol_type="genomic DNA"
         /db_xref="taxon:9606"
         /chromosome="9"
         /clone="RP11-108L4"
         /clone_1fb="RP11-11.1"
      184..221
         /note="MIR repeat: matches 36..79 of consensus"
      1041..1323
         /note="AluSP repeat: matches 1..284 of consensus"
      1827..1997
         /note="MLT1I repeat: matches 13..182 of consensus"
      2533..2757
         /note="MER46A repeat: matches 3..236 of consensus"
      3466..3773
         /note="AluSQ repeat: matches 1..308 of consensus"
      4020..4059
         /note="MLT1I repeat: matches 330..370 of consensus"
      4243..4506
         /note="AluSQ repeat: matches 1..264 of consensus"
      4517..4813
         /note="L1M4 repeat: matches 2336..2645 of consensus"
      5027..5158
         /note="L1M4 repeat: matches 2957..3090 of consensus"
      6903..6990
         /note="L2 repeat: matches 1756..1843 of consensus"
      6991..7479
         /note="MLT1F repeat: matches 1..507 of consensus"
      7480..7776
         /note="AluDb repeat: matches 1..305 of consensus"
      7777..7813
         /note="MLT1F repeat: matches 507..541 of consensus"
      8558..8613
         /note="L2 repeat: matches 2685..2741 of consensus"
      8717..8827
         /note="AluSP repeat: matches 2..86 of consensus"
      8829..9169
         /note="LTR13 repeat: matches 28..364 of consensus"
      9167..9808
         /note="LTR13 repeat: matches 308..945 of consensus"
      9809..10024
         /note="AluSP repeat: matches 81..297 of consensus"
      10719..10949
         /note="MER58A repeat: matches 1..222 of consensus"
      11780..12114
         /note="MLT1B repeat: matches 1..376 of consensus"
      12271..12583
         /note="AluSQ repeat: matches 1..313 of consensus"
      13003..13783
         /note="L1PA2 repeat: matches 2..776 of consensus"
      13779..15033
         /note="L1PA2 repeat: matches 900..6146 of consensus"
      19133..19209
         /note="MER30 repeat: matches 136..230 of consensus"
      20138..20243
         /note="53 copies 2 mer tt 62% conserved"
      20399..20433
         /note="MIR repeat: matches 112..146 of consensus"
      21052..21340
         /note="AluSQ repeat: matches 1..310 of consensus"
      21375..21414
         /note="20 copies 2 mer aa 77% conserved"

```

repeat_region 21512..21819
/note="AluY repeat: matches 1..309 of consensus"
repeat_region 21979..22174
/note="MIR repeat: matches 29..227 of consensus"
repeat_region 22282..22467
/note="LIMB3 repeat: matches 4603..4791 of consensus"
repeat_region 22470..22629
/note="FRAM repeat: matches -1..164 of consensus"
repeat_region 22630..23631
/note="LIMB3 repeat: matches 4786..5736 of consensus"
repeat_region 23632..23994
/note="WSTA repeat: matches 1..425 of consensus"
repeat_region 23995..24423
/note="LIMB3 repeat: matches 5736..6185 of consensus"
repeat_region 25327..25632
/note="AluSP repeat: matches 1..299 of consensus"
repeat_region 25647..25935
/note="LIMB4 repeat: matches 6005..6300 of consensus"
repeat_region 26753..26792
/note="20 copies 2 mer at 80% conserved"
repeat_region 26805..26852
/note="24 copies 2 mer ta 77% conserved"
repeat_region 26946..27277
/note="LIMB7 repeat: matches 5933..6260 of consensus"
repeat_region 27721..28015
/note="AluX repeat: matches 1..295 of consensus"
repeat_region 28379..28761
/note="MIRB repeat: matches 1..388 of consensus"
repeat_region 29408..29719
/note="Aluub repeat: matches 1..312 of consensus"
repeat_region 29737..29776
/note="FLAM repeat: matches 86..125 of consensus"
repeat_region 29777..30004
/note="Aluud repeat: matches 1..220 of consensus"
repeat_region 32322..32617
/note="Aluub repeat: matches 6..298 of consensus"
repeat_region 32656..32766
/note="MIR repeat: matches 23..139 of consensus"
repeat_region 32996..33035
/note="Alu repeat: matches 66..105 of consensus"
repeat_region 35794..35944
/note="AluSc repeat: matches 133..277 of consensus"
repeat_region 36404..36705
/note="AluSg repeat: matches 1..302 of consensus"
repeat_region 37126..37446
/note="AluX repeat: matches 1..311 of consensus"
repeat_region 38017..38301
/note="AluX repeat: matches 1..285 of consensus"
repeat_region 38687..38914
/note="L2 repeat: matches 2532..2746 of consensus"
repeat_region 38917..39095
/note="MER20 repeat: matches 34..218 of consensus"
repeat_region 39424..39689
/note="AluSg repeat: matches 1..266 of consensus"
repeat_region 41016..41309
/note="AluX repeat: matches 2..297 of consensus"
repeat_region 42261..42665
/note="MIR2F repeat: matches 1..399 of consensus"
repeat_region 43512..43653
/note="ORSL repeat: matches 218..358 of consensus"
repeat_region 44023..44196
/note="MERSA repeat: matches 3..189 of consensus"
repeat_region 45028..45321
/note="AluX repeat: matches 1..296 of consensus"
repeat_region 45473..45586
/note="L2 repeat: matches 2630..2750 of consensus"
repeat_region 45785..45990
/note="MIR repeat: matches 21..255 of consensus"
repeat_region 47204..47488
/note="AluX repeat: matches 1..285 of consensus"
repeat_region 55108..55341
/note="MIR repeat: matches 27..260 of consensus"
repeat_region 57364..57658

repeat_region /note="Aluub repeat: matches 1..295 of consensus"
57748..57811
/note="L2 repeat: matches 2646..2709 of consensus"
repeat_region /note="L2 repeat: matches 2646..2709 of consensus"
59230..59375
/note="MERSA repeat: matches 18..188 of consensus"
repeat_region 59583..59654
/note="36 copies 2 mer tt 65% conserved"
repeat_region 59658..59977
/note="LIPB7 repeat: matches 5802..6141 of consensus"
repeat_region 61302..61714
/note="L2 repeat: matches 1248..1673 of consensus"
repeat_region 61785..62178
/note="L2 repeat: matches 1827..2199 of consensus"

Query Match 100.0%; Score 228; DB 9; Length 128480;
Best Local Similarity 100.0%; Pred. No. 1..1e-114;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTAAGTCCTTATTAAGACTCAATTTCTACTGAGAAATTTTGATGGCTTAAGTTCT 60
Db 126852 GTAAGTCCTTATTAAGACTCAATTTCTACTGAGAAATTTTGATGGCTTAAGTTCT 126911

QY 61 CTACTGCTTTCTATCTCTCTACTGACTGCTCCCGGATCCACTACGATTTTCTATT 120
Db 126912 CTACTGCTTTCTATCTCTCTACTGACTGCTCCCGGATCCACTACGATTTTCTATT 126971

QY 121 TCTTGCTGATTTGCTGACTGCTGCTGATTTATCTTACGAGCTGATTTTC 180
Db 126972 TCTTGCTGATTTGCTGACTGCTGCTGATTTATCTTACGAGCTGATTTTC 127031

QY 181 TACCCGGCTACCTCCGCTCCATATTTGCTCCACTTTACAG 228
Db 127032 TACCCGGCTACCTCCGCTCCATATTTGCTCCACTTTACAG 127079

RESULT 3
AF103908
LOCUS Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
DEFINITION AF103908
ACCESSION AF103908.1 GI:6165974
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 5435)
Bussemakers,M.J., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Kathaus,H.F., Schalken,J.A., Debruyne,F.M., Ru.N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate cancer
Cancer Res. 59 (23), 5975-5979 (1999)
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 5435)
10606244
Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Kathaus,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru.N. and
Isaacs,W.B.
Direct Submission
Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
LOCATION/Qualifiers
1..5435
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..78
/rpt_family="AluY"
/rpt_type="dispersed"
533..697
/note="alternative exon present in 5% of cDNA clones"

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preerred.

1 41864: contig of 41864 bp in length
* 41865 41964: gap of 100 bp
* 41965 42641: contig of 677 bp in length
* 42642 42741: gap of 100 bp
* 42742 43862: contig of 1121 bp in length
* 43863 43962: gap of 100 bp
* 43963 44630: contig of 668 bp in length
* 44631 44730: gap of 100 bp
* 44731 46031: contig of 1301 bp in length
* 46032 46131: gap of 100 bp
* 46132 48459: contig of 2328 bp in length
* 48460 48559: gap of 100 bp
* 48560 52262: contig of 3703 bp in length
* 52263 52362: gap of 100 bp
* 52363 55291: contig of 2929 bp in length
* 55292 55391: gap of 100 bp
* 55392 57454: contig of 2063 bp in length
* 57455 59456: gap of 1902 bp in length
* 59457 59556: gap of 100 bp
* 59557 62557: contig of 3001 bp in length
* 62558 62657: gap of 100 bp
* 62658 65408: contig of 2751 bp in length
* 65409 65508: gap of 100 bp
* 65509 69990: contig of 4482 bp in length
* 69991 70090: gap of 100 bp
* 70091 73982: contig of 3892 bp in length
* 73983 74082: gap of 100 bp
* 74083 80679: contig of 6597 bp in length
* 80680 80779: gap of 100 bp
* 80780 94247: contig of 13468 bp in length
* 94248 94347: gap of 100 bp
* 94348 108642: contig of 14295 bp in length
* 108643 108742: gap of 100 bp
* 108743 125926: contig of 17184 bp in length
* 125927 126026: gap of 100 bp
* 126027 145344: contig of 19318 bp in length
* 145345 145444: gap of 100 bp
* 145445 166679: contig of 21235 bp in length
* 166680 166780: gap of 100 bp
* 166780 224052: contig of 57273 bp in length.

FEATURES

source
1..224052
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-222013"
/clone_lib="RPCT-24 Male Mouse BAC"
1..41864
/note="assembly_fragment"
clone_end:SP6
vector_side:left
41965..42641
/note="assembly_fragment"
42742..43862
/note="assembly_fragment"
43963..44630
/note="assembly_fragment"
44731..46031
/note="assembly_fragment"
46132..48459
/note="assembly_fragment"
48560..52262
/note="assembly_fragment"
52363..55291
/note="assembly_fragment"
55392..57454
/note="assembly_fragment"
57555..59456
/note="assembly_fragment"

misc_feature 59557..62557
/note="assembly_fragment"
misc_feature 62658..65408
/note="assembly_fragment"
misc_feature 65509..69990
/note="assembly_fragment"
misc_feature 70091..73982
/note="assembly_fragment"
misc_feature 74083..80679
/note="assembly_fragment"
misc_feature 80780..94247
/note="assembly_fragment"
misc_feature 94348..108642
/note="assembly_fragment"
misc_feature 108743..125926
/note="assembly_fragment"
misc_feature 126027..145344
/note="assembly_fragment"
misc_feature 145445..166679
/note="assembly_fragment"
misc_feature 166780..224052
/note="assembly_fragment"

BASE COUNT 58760 a 51676 c 52510 g 59105 t 2001 others
ORIGIN

Query Match 10.1%; Score 23; DB 2; Length 224052;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dy 74 TCCTTCTACTCACTGCTCTCC 96
Db 81046 TCCTTCTACTCACTGCTCTCC 81068

RESULT 5
AC115779/c 198517 bp DNA linear HTG 27-MAR-2003
LOCUS
DEFINITION Mus musculus clone RP23-153122, WORKING DRAFT SEQUENCE, 12
unorderd pieces.

ACCESSION AC115779
VERSION AC115779.5 GI:29294112
KEYWORDS HTG; HTGS PHASE; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 198517)
REFERENCE
1 Birren, B., Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalil, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, C.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menue, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 198517)
Bairren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., DeRellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Hagagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, D., Menus, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retter, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 27, 2003 this sequence version replaced gi:27764772.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 123444
Center clone name: 153_L_22
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 195514 bases at least Q40
Consensus quality: 196343 bases at least Q30
Consensus quality: 196640 bases at least Q20
Insert size: 197417; sum-of-coverage
Quality coverage: 10.1 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 17349: contig of 17349 bp in length
* 17350
* 17449: gap of 100 bp
* 17450
* 18521: contig of 1072 bp in length
* 18522
* 18621: gap of 100 bp
* 18622
* 18626: contig of 905 bp in length
* 18627
* 19626: gap of 100 bp
* 19627
* 21227: contig of 1601 bp in length
* 21228
* 21327: gap of 100 bp
* 21328
* 24152: contig of 2825 bp in length
* 24153
* 24252: gap of 100 bp
* 24253
* 29267: contig of 5515 bp in length
* 29268
* 29867: gap of 100 bp
* 29868
* 35683: contig of 5816 bp in length
* 35684
* 35783: gap of 100 bp

FEATURES
source
1. 198517
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-153L22"
/clone_1b="RPCI-23 Female Mouse BAC"
1. 17349
/note="assembly_fragment"
clone_end:SP6
vector_side:left
17450..18521
/note="assembly_fragment"
18622..19526
/note="assembly_fragment"
19627..21227
/note="assembly_fragment"
21328..24152
/note="assembly_fragment"
24253..29267
/note="assembly_fragment"
29868..35683
/note="assembly_fragment"
35784..46552
/note="assembly_fragment"
4653..68280
/note="assembly_fragment"
68381..87560
/note="assembly_fragment"
87661..166311
/note="assembly_fragment"
166312
166312
198517: contig of 32106 bp in length.
location/Qualifiers

misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:left
17450..18521
/note="assembly_fragment"
18622..19526
/note="assembly_fragment"
19627..21227
/note="assembly_fragment"
21328..24152
/note="assembly_fragment"
24253..29267
/note="assembly_fragment"
29868..35683
/note="assembly_fragment"
35784..46552
/note="assembly_fragment"
4653..68280
/note="assembly_fragment"
68381..87560
/note="assembly_fragment"
87661..166311
/note="assembly_fragment"
166412..198517
/note="assembly_fragment"
clone_end:T7
vector_side:right

BASE COUNT 59669 a 38866 c 39210 g 59671 t 1101 others
ORIGIN

Query Match 9.6%; Score 22; DB 2; Length 198517;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATAAAGCATCAATTTCTACTC 33
Db 181610 ATAAAGCATCAATTTCTACTC 181589

RESULT 6
AL929582 175577 bp DNA linear VRT 23-APR-2003
LOCUS Zebrafish DNA sequence from clone CH211-117M12, complete sequence.
ACCESSION AL929582
VERSION AL929582.10 GI:30024470
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 175577)
AUTHORS Dunn, M.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

COMMENT

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 18, 2003 this sequence version replaced gi:29823207.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rexio/fishmask.shtml

VECTOR: pTARBAC2.1.
Location/Qualifiers

FEATURES

source

1. 175577
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_id="CH211-117M12"
/clone_lib="CHORI-211"

BASE COUNT 54854 a 33089 c 32287 g 55347 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 1.9; Length 175577;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ATTCTTGCTCGATTGTCT 138

Db 154212 ATTCTTGCTCGATTGTCT 154232

RESULT 7
AL935269/c 177658 bp DNA linear VRT 11-APR-2003

LOCUS Zebrafish DNA sequence from clone DKEY-46J18, complete sequence.

ACCESSION

AL935269

VERSION

AL935269.7

KEYWORDS

HTG.

SOURCE

ORGANISM

Danio rerio (zebrafish)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rexio/fishmask.shtml

DKEY-46J18 is from a Zebrafish BAC library

FEATURES

source

1. 177658
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_id="DKEY-46J18"
/clone_lib="DanioKey"

BASE COUNT 56507 a 31931 c 32628 g 56592 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 1.8; Length 177658;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ATTCTTGCTCGATTGTCT 138

Db 27032 ATTCTTGCTCGATTGTCT 27012

RESULT 8
AX107860/c 20 bp DNA linear PAT 30-APR-2001

LOCUS Sequence 4 from Patent WO0123550.

ACCESSION

AX107860

VERSION

AX107860.1

KEYWORDS

SOURCE

ORGANISM

synthetic construct

artificial sequences.

REFERENCE

AUTHORS

TITLE

JOURNAL

Diagnocure Inc. (CA)

Patent: WO 0123550-A 4 05-APR-2001;

Diagnocure Inc. (CA)

Diagnocure Inc. (CA)

Diagnocure Inc. (CA)

Diagnocure Inc. (CA)

Diagnocure Inc. (CA)

Diagnocure Inc. (CA)

Diagnocure Inc. (CA)

Query Match 8.8%; Score 20; DB 6; Length 20;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 66 CGTTCTATCTCTCTACTC 85
 |||
 Db 20 CGTTCTATCTCTCTACTC 1

RESULT 9
 AC087595/c 153479 bp DNA linear HTG 13-JAN-2001
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 1 clone
 DEFINITION OSUNBA0034K07, ** SEQUENCING IN PROGRESS **, 2 ordered pieces.
 ACCESSION AC087595.1 GI:12203307
 VERSION HTG; HTGS PHASE2.
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 153479)
 Lee,J.-S., Hahn,J.-H., Lee,M.-C., Yoon,U.-H., Yun,D.-W., Kim,H.-I.
 and Eun,M.Y.
 Oryza sativa BAC OSUNBA0034K07 genomic sequence
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 153479)
 REFERENCE Hahn,J.-H. and Eun,M.Y.
 AUTHORS Direct Submission
 JOURNAL Submitted (13-JAN-2001) Rice Genome Sequencing Project, National
 Institute of Agricultural Science and Technology (NIIST), RDA, 249
 Seodun-dong, Suwon 441-707, Korea (E-mail:jhahn@rda.go.kr,
 Tel:82-31-290-0309, Fax:82-31-290-0308)

COMMENT * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 45287: contig of 45287 bp in length
 * 45288 45387: gap of unknown length
 * 45388 153479: contig of 108092 bp in length.

FEATURES
 source
 1.153479
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="1"
 /clone="OSUNBA0034K07"
 BASE COUNT 44369 a 33690 c 32959 g 42321 t 100 others
 ORIGIN

Query Match 8.8%; Score 20; DB 2; Length 153479;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 102 CCAGTACGATTTTCTATT 121
 |||
 Db 117052 CCAGTACGATTTTCTATT 117033

RESULT 10
 AP003451/c 162161 bp DNA linear PLN 29-MAR-2003
 LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 DEFINITION PAC clone:P0413C03.
 ACCESSION AP003451 BA000010
 VERSION AP003451.4 GI:19386744

KEYWORDS
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
 Katayose,Y., Wu,J., Nishimura,Y., Cheng,Z., Nagamura,Y.,
 Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Aikawa,K.,
 Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
 Hamada,M., Harada,C., Hijiwata,S., Honda,M., Ichikawa,Y.,
 Iidoma,A., Iijima,M., Ikeda,M., Ikono,M., Itoh,S., Itoh,T.,
 Itoh,Y., Itoh,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
 Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,
 Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
 Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
 Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
 Shomura,A., Song,U., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
 Yamagata,H., Yamane,H., Yoshiki,S., Yoshinara,R., Yukawa,K.,
 Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.-H., Kim,H.-I., Eun,M.Y.,
 Yano,M., Jiang,J. and Gojobori,T.
 The genome sequence and structure of rice chromosome 1
 Nature 420 (6913), 312-316 (2002)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 AUTHORS Direct Submission
 JOURNAL Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program; Kamondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/
 Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT On Mar 12, 2002 this sequence version replaced gi:16904688.
 Genes were predicted from the integrated results of the following:
 GENSCAN1.0, BLASTX2.0, BLASTX2.0 as well as SplicePredictor
 (October 1998 version). The genomic sequence was searched against
 NCBI Nonredundant Protein database, nr
 (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
 RGP. Protein homologies of the coding regions were searched against
 NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
 the identified cDNA sequences using BLASTN 2.0 with the
 corresponding DDBJ accession no. and RGP clone ID.
 A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without
 significant homology to any protein but with EST homology (covering
 almost the entire length of partial sequence) is classified as an
 'unknown' protein. A gene predicted with a gene prediction program
 is classified as a 'hypothetical' protein.
 The orientation of the sequence is from SP6 to T7 of the PAC clone.
 This sequence of P0413C03 clone has an overlap with P0678F11 (DDBJ:
 AB003437) clone at the position 148,366 to 150,019 of 3' end. The
 sequence of this clone ends at the position 13,793 of P0678F11.
 Detailed information on overlap and assembly quality together with
 annotation of this entry is available at
 http://rsgp.dna.affrc.go.jp/Genomeseq.html.

FEATURES
 source
 1.162161
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="1"
 /clone="P0413C03"
 join(1042..1208,1804..1858)
 /gene="P0413C03.1"
 join(1042..1208,1804..1858)
 /gene="P0413C03.1"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB86126.1"

/db_xref="GI:19386745"
/translation="MRGDGDEGCADTQDEGRDNDGGDGADEBPILLTLPHSL
LQYTLIPVDLTCRSTRYIMPNTRKLDK"
complement (4165..5010)
/gene="P0413C03.2"
complement (4165..5010)
/note="hypothetical protein"
/protein_id="BAB86127.1"
/db_xref="GI:19386746"
/translation="MSSWNSPYYTSSYGAGSGGGGGRSSPPRAGCKYTK
DYQYTSNNNSDDDKDKNKHKITSHRKDEKRNHNSKDSHGSGNSNNKKS
YGGSGNPNYGGSTGVAGSGSYGGGAGGANTSYGSLSYGKDGSGNSPY
GGSSIIISGAAPIPHNFGGCTGVPYPPDGGSGAPVFIKAEVKIYHTTP
GHGSSSSSPSPKYESBGRDRRSAGAGSGSGSGGCGGFFGPAPFAVGXIDRKF
GLDKD"
complement (7475..8275)
/gene="P0413C03.3"
complement (7475..8275)
/gene="P0413C03.3"
/note="contains ESTs AU10652 (C11143), C25923 (C11143)
unknown protein"
/codon_start=1
/protein_id="BAB86128.1"
/db_xref="GI:19386747"
/translation="MSWKKSSGSSSSQYVAGSDSSFPAGGGDGRYGRAGSGNG
RSPWTSSEYRYTSDNDRSGSDDDDDHNRNRKKNKSSSSSSSKYD
DGBRDRSYNYGYGNNEGYSSAPSYGNNYNGGYSYNGCGYGGSTYNGG
GYGSSYNGGCAIVAGGGGAPASYSYGSNGAPAPPRGEGSGAPYLVNIT
TOPSSSSQNSERDNGGGGNGLFGPTQAVGVMDRRFFD"
complement (10814..11478,13318..13418,14749..14934,
15151..15224)
/gene="P0413C03.4"
complement (10814..11478,13318..13418,14749..14934,
15151..15224)
/gene="P0413C03.4"
/note="contains EST AU068098 (C11856)
unknown protein"
/codon_start=1
/protein_id="BAB86129.1"
/db_xref="GI:19386748"
/translation="MEMPFYGVSLVYARITAGVOVSYTSLTTHSIGOMSFHD
TNITFYTHPIKILKYLIMPTADNNLKGVIKWRYLKRERERLPKTCR
FNACVSEKSTALVLDPSGMNNSGSPSSGSSNDTSGGROOQOQOQYEA
AAGDKTAKQPPSPRLKTKSGNDDAATAVGVYASGGYGGSTKANAGYGG
GGYNGCYNRAGSTLYGAGVGTPTYYGGGGGGTGGGSPVYGGGVPAGYTPONG
ARSPLYNTRVHYGHPYDGDGDDGGGSGGERRRSGFFRPAFAVGHFDR
FGHSSD"
complement (16470..17480,17716..17859,18651..18728,
19173..19289,19376..19432,19943..20074,20176..20373,
20815..20925,21021..21236,21330..21515,22041..22643)
/gene="P0413C03.5"
complement (16470..17480,17716..17859,18651..18728,
19173..19289,19376..19432,19943..20074,20176..20373,
20815..20925,21021..21236,21330..21515,22041..22643)
/gene="P0413C03.5"
/codon_start=1
/product="putative gamma-adaptin 1"
/protein_id="BAB86130.1"
/db_xref="GI:19386749"
/translation="MEOLRTIGRELAMGSGGMSKEFLDLVKSIGEASKAEEDRI
ISRLDLKRLADPDVPRRMKELLRLVLAELGHDASGHTHAYKMDDELPIK
RTGYLAFLDDEPHDLVILYNTQIDRLSDNTLVYCAALTAACRIGEAATYAVP
QVVELLHPKEAVRKAVMALHRYQSPSSVSLVSNFRRLCDNPGWGAATLCEL
YDLLEDPNSTKDLVSVNLTQVARRLPSTVDYQMPAPFQILKTLAVLGG
DKQASGNTKMGVIGLIRKIDINPIAEHQAVIDCEDPDITKRFELLYGMS
KSQSHNLKMGVIGLIRKIDINPIAEHQAVIDCEDPDITKRFELLYGMS
TNVEVIVDMLEWYINITDHYKTEISRCCELAEOAPASQWOTIOMNRYEYHGS
VNIRVANNMLIAEGGERDEGASOLRSASVSYRICEPTELLBSFLOITWGL
EYGRADKYSASYIKLVDVAEHPDPTDTRAYTAISLIKTFEFLALGKIDMLES
COSLIDELASHSHTDQRAVELQALGLDQAEVNWADSCEDIIDRNLSFLNG
YVQALENGAPVPIESERSGVSVGVYKADQDOETSAHALRFAVELPRAASQASIS

PTDVLVPEPSPYKEDHQMSRQPSGDSLSGEFLKRLDGVQKMGAPYSSSTP
SSSTSOQATNGVSSSEVGSTSGQAESTRYGSROQATEWSAEKHLASLFEKADR
KTOAGRTAKESSTTEKATVATNPQAKGVIPASPPDLIDGESSSPPLADPF
TOLEGLGASASPVSETPASTSTKPDLSIFSPDVQCVTSGTEBPLGVNVAA
KKGPSLQDALQKMAAQVAVTPGNRPILFKDLG"
complement (124745..24946,25096..25194,25490..25709,
26050..26215,26559..26612,26978..27108,27228..27269,
27975..28053,28156..28234,28437..28597)
/gene="P0413C03.6"
complement (124745..24946,25096..25194,25490..25709,
26050..26215,26559..26612,26978..27108,27228..27269,
27975..28053,28156..28234,28437..28597)
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome 2, At2g1020"
/codon_start=1
/protein_id="BAB86131.1"
/db_xref="GI:19386750"
/translation="MQRATLLHNPPCLSLTWMLGYNQIRQLMKPPTIMIVOQTNL
LSKRLFPVNRQAKATIEPTQYEDLLSNRNPALKHEILKMTAEHRAEMANKRKEPL
HADGNVEIGNGVPGGDKDIADAKADDLPEFLRRLRGLIKDQATNRRFT
IKONVDSPVQIYKAAQELPGWYEAQKPTGASVYFVQSGTQWDRPGAPLNMQCQ
APSSSLPENMEALDQSTQKYYTNTNTATQMEPPTAVNAPVAPPAVNAVEMKA
QNTDINMSQRCICGCGWGLVQPWGYNHCTRVQLPFOQYPSYNNTHSSNK
NPGVNAKDRSSAKPPFGKANRKHRRKRRPEDELDPMDPSSYSIDAPRGWYIFPI
LVNYASFEIDYLPVSW"
/join (28678..28714,29236..29459,29480..29569)
/gene="P0413C03.7"
/join (28678..28714,29236..29459,29480..29569)
/gene="P0413C03.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB86132.1"
/db_xref="GI:19386751"
/translation="MEGGGHAFTPYGYRYPAGGSPFPAAGRPICGVARLDKGE
EPQERAPSELSCGGSVDCRRRNPRLRRGGAPSGAPAPPAATPVPL"
/join (32933..33059,33753..33948,34052..34279,34608..34839,
34927..35034,35147..35419,35712..35996)
/gene="P0413C03.8"
/join (32933..33059,33753..33948,34052..34279,34608..34839,
34927..35034,35147..35419,35712..35996)
/gene="P0413C03.8"
/note="contains ESTs
AU065676 (E50933), D24223 (R1545), AU030352 (E50933)"
/codon_start=1
/product="putative selenium binding protein"
/protein_id="BAB86133.1"
/db_xref="GI:19386752"
/translation="MAAANGAACCGGATPGYATPLEAMEKGFREKTLVTVYVNGT

Query Match 8.8%; Score 20; DB 8; Length 16216;
Beet Local Similarity 100.0%; Pred. No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 102 CCACTACCGATTTTCTATT 121
Db 132589 CCACTACCGATTTTCTATT 132570

RESULT 11
AC122812
LOCUS AC122812 208430 bp DNA linear ROD 06-SEP-2002
DEFINITION Mus musculus chromosome 1 clone RP28-285M9, complete sequence.
ACCESSION AC122812
VERSION AC122812.4 GI:22748565
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 208430)
MCPHERSON, J.D. and WATERSTON, R.H.
AUTHORS The sequence of Mus musculus clone

JOURNAL
REFERENCE 2 (bases 1 to 208430)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 208430)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-ANG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 208430)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Sep 6, 2002 this sequence version replaced gi:22475552.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BA0285W09

FEATURES
source
1. location/Qualifiers
1.208430
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP23-285W9"
BASE COUNT 61643 a 40063 c 41477 g 65247 t
ORIGIN

Query Match 8.8%; Score 20; DB 10; Length 208430;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 106 TACCGATTTCTATTCTTG 125
|||||
Db 93902 TACCGATTTCTATTCTTG 93921

RESULT 12
AC109025
LOCUS
DEFINITION AC109025 215616 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-276A24, *** SEQUENCING IN PROGRESS

AC109025
VERSION AC109025.5 GI:25139227
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 215616)
Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrechts, S., Amin, A., Angiano, D.,
Ayala, Bechl, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryan, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, J., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
DeGado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., DiVya, K.,
Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., DiVya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flag, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabrial, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladik, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howell, S., Huyl, S., Hume, J., Idelbird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuwa, L., Louieged, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Mloaavilev, A., Miner, G., Ming, A., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Nwokedeme, O., Okunolu, G., Olarnpusagoon, A., Pal, S.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,
Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Riley, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valae, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, K., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, Y.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Miederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinreich, G., and Gibbs, R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215616)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 215616)
REFERENCE Rat Genome Sequencing Consortium.
JOURNAL Direct Submission
JOURNAL Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT The sequence in this assembly is a combination of BAC based reads
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GPRH
Center clone name: CH230-276A24
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 181529 bases at least Q40


```

misc_feature      /note="assembly_name:Contig23"
                  59207..88462
                  /note="assembly_name:Contig24"
misc_feature      88563..134957
                  /note="assembly_name:Contig25"
misc_feature      135058..170399
                  /note="assembly_name:Contig26"
misc_feature      170500..230279
                  /note="assembly_name:Contig27"
BASE COUNT      63548 a 53964 c 52004 g 59749 t 1014 others
ORIGIN
Query Match      8.8%; Score 20; DB 2; Length 230279;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AAGTCTTTATTAAGCACTC 22
Db      204760 AAGTCTTTATTAAGCACTC 204779

RESULT 14
AC141860/c      AC141860      237234 bp      DNA      linear      HTG 03-APR-2003
LOCUS      Gallus gallus clone TAM31-57D16, WORKING DRAFT SEQUENCE, 10 ordered
DEFINITION      pieces.
AC141860
VERSION      AC141860.2 GI:29501863
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 237234)
Ahter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakeley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Cariaga,C., Coleman,B., Engle,J., Grant,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Ido,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., McDowell,J.,
Murgitane,C., Pearson,R., Portnoy,M.E., Prasad,A.,
Reddix-Dugue,N., Schandler,K., Schuler,M.G., Sison,C.,
Santirpoo,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Weheby,K.D., Wiggins,L., Young,A. and Green,E.D.
NSC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 237234)
Green,E.D.
Direct Submission
Submitted (19-MAR-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 237234)
Green,E.D.
Direct Submission
Submitted (03-APR-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
On Apr 3, 2003 this sequence version replaced gi:29124110.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@hgrl.nih.gov
----- Project Information
Center project name: dft
Center clone name: 057D16

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g.,

```

human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990139
Consensus quality: 235153 bases at least Q40
Consensus quality: 235850 bases at least Q30
Consensus quality: 236231 bases at least Q20
Insert size: 180000; agarose-efp
Insert size: 236334; sum-of-contigs
Quality coverage: 11.59x in Q20 bases; agarose-efp
Quality coverage: 8.83x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 12817: contig of 12817 bp in length
*
12818 12917: gap of unknown length
*
12918 85073: contig of 72156 bp in length
*
85074 85173: gap of unknown length
*
85174 88714: contig of 3541 bp in length
*
88715 88815: gap of unknown length
*
88815 113928: contig of 25114 bp in length
*
113929 114028: gap of unknown length
*
114029 191667: contig of 77639 bp in length
*
191668 200289: contig of 8522 bp in length
*
200290 200389: gap of unknown length
*
200390 211976: contig of 11557 bp in length
*
211977 212076: gap of unknown length
*
212077 214145: contig of 2069 bp in length
*
214146 214245: gap of unknown length
*
214246 224194: contig of 9949 bp in length
*
224195 224294: gap of unknown length
*
224295 237234: contig of 12940 bp in length.
*
FEATURES
Source
1..237234
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/clone="TAM31-57D16"
/clone_1b="TAM31"
1..12817
/note="assembly_fragment
clone_end:T7
vector_side:left"
12818..85073
/note="assembly_fragment"
85174..88714
/note="assembly_fragment"
88815..113928
/note="assembly_fragment"
114029..191667
/note="assembly_fragment"
191768..200289
/note="assembly_fragment"
200390..211976
/note="assembly_fragment"
212077..214145
/note="assembly_fragment"
214246..224194
/note="assembly_fragment"

```



```

misc_feature      224295..237234
                  /note="assembly_fragment
                  /clone_end:SP6
                  vector_side:right"
BASE COUNT      69220 a 46694 c 48269 g 72151 t      900 others
ORIGIN
Query Match      8.8%; Score 20; DB 2; Length 237234;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      67 GTTTCTAATCTCTCTACTCA 86
          |||||
Db      227181 GTTTCTAATCTCTCTACTCA 227162

RESULT 15
AC134221/c
LOCUS
DEFINITION      AC134221
                  Rattus norvegicus clone CH230-1005, WORKING DRAFT SEQUENCE, 2
                  unordered pieces.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 238817)
Muzny,D,Marle,M, Metzker,M, Lee,A, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alshrooke,S, Amin,A, Angiano,D,
Ayala-Bechechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryan,N, Buhay,C, Burch,P, Butrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Devita,M, Davis,C, Day-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
Diaper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Gebreyes,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guetara,W,
Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,B, Howell,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovat,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorensuwa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E,
Mawhinley,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Naokoleheneh,O, Okunonu,G, Olarnunagoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C,
Plopper,F, Poldexker,A, Popovic,D, Prymne,B, Pu,L,L,
Piazzi,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
Reilly,M, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Rivers,C, Rodkey,T, Rojars,A, Rose,M, Rose,R, Ruiz,S,J,
Sanders,W, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,
Shetty,J, Shvartsbeyn,A, Sison,I, Sitter,C,D, Smajic,D,
Sneed,A, Sodergren,E, Song,X,Z, Sorelle,R, Sosa,J,
Steinle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C,
Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K,
Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Wainstock,G., and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 238817)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (24-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 238817)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23306103.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDUS
Center clone name: CH230-1005
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 221775 bases at least Q40
Consensus quality: 224525 bases at least Q30
Consensus quality: 226115 bases at least Q20
Estimated insert size: 226851; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 135250: contig of 135250 bp in length
* 135251 135350: gap of unknown length
* 135351 238817: contig of 103467 bp in length.
Location/Qualifiers
1..238817
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-1005"
87008..89104
/note="wgs_contig"
135351..136555
/note="wgs_contig"
139028..141263
misc_feature
misc_feature
misc_feature
misc_feature

```

```

BASE COUNT      65338 a 49322 c 49101 g 63267 t 11789 others
ORIGIN
Query Match      8.8%; Score 20; DB 2; Length 238817;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      198 GTCCCTCCATTGTCCTC 217
Db      36613 GTCCCTCCATTGTCCTC 36594

RESULT 16
AC097028/c      243174 bp      DNA      linear      HTG 10-MAY-2003
LOCUS
DEFINITION      Rattus norvegicus clone CH230-93J21, WORKING DRAFT SEQUENCE.
ACCESSION
AC097028
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 243174)
Mizny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Denson,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falle,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnarsson,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huliy,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenzhuwa,L., Loulseged,H., Lorzado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemelehen,O., Okwomu,G., Olariunbasagun,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,A., Pflanzko,C.,
Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Riley,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shateman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smaje,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,R.,
Steimle,M., Strong,R., Sutton,A., Swatek,A., Taber,Z., Tamari,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemari,K.,
Valas,R., Vera,Y., Villaseca,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Woodson,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Niederhausern,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 243174)
Worley,K.C.
Submitted Submission
Submitted (06-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 243174)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:2265500.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G3XU
Center clone name: CH230-93J21
----- Summary Statistics
Assembly program: Atlas 3.0:
Consensus quality: 234402 bases at least Q40
Consensus quality: 236580 bases at least Q30
Consensus quality: 237637 bases at least Q20
Estimated insert size: 243918; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 243174: contig of 243174 bp in length.
Location/Qualifiers
1. 243174
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-93J21"
213749..215121
misc_feature
/note="wgs_end_extension"
clone_end:T7
233065..233105
misc_feature
/note="clone_boundary"
clone_end:T7
site:ECORI
end_sequence:BH327070"
complement(240473..241336)
/note="clone_boundary"

```

```

clone end:Sp6
site:ĖcorI
end sequence: BH327071"
BASE COUNT      66968 a 53236 c 53017 g 64871 t 5082 others
ORIGIN

```

	a	c	g	t	others
BASE COUNT ORIGIN	66968	53236	53017	64871	5082

ORIGIN

Query Match	8.8%	Score 20	DB 2	Length 243174
Best Local Similarity	100.0%	Pred. No. 6	4	
Matches 20	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	198	GTCCTTCATATTGTCCTC	217
Db	25449	GTCCTTCATATTGTCCTC	25430

RESULT 17			
AC122189			
LOCUS	AC122189	247309 bp	DNA
DEFINITION	Mus musculus chromosome 1 clone RP23-36c8, complete sequence.	linear	ROD 17-OCT-2002

SOURCE	ORGANISM
Mus musculus (house mouse)	Mus musculus
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia, Eutheria, Rodentia; Sciurognathi; Muridae; Murinae; Mus	

REFERENCE	JOURNAL	McPherson,J.D. and Waterston,R.H.	4444 Forest Park
AUTHORS	TITLE	The Sequence of Mus musculus clone	
AUTHORS	TITLE	Unpublished	
REFERENCE	JOURNAL	2 (bases 1 to 247309)	
AUTHORS	TITLE	McPherson,J.D. and Waterston,R.H.	
AUTHORS	TITLE	Direct Submission	
JOURNAL	TITLE	Submitted (23-May-2002) Genome Sequencing Center,	4444 Forest Park
REFERENCE	JOURNAL	Parkway, St. Louis, MO 63108, USA	
AUTHORS	TITLE	3 (bases 1 to 247309)	
AUTHORS	TITLE	McPherson,J.D. and Waterston,R.H.	
JOURNAL	TITLE	Direct Submission	
REFERENCE	JOURNAL	Submitted (10-Oct-2002) Genome Sequencing Center,	4444 Forest Park
AUTHORS	TITLE	Parkway, St. Louis, MO 63108, USA	
AUTHORS	TITLE	4 (bases 1 to 247309)	
JOURNAL	TITLE	McPherson,J.D. and Waterston,R.H.	
REFERENCE	JOURNAL	Direct Submission	
AUTHORS	TITLE	Submitted (17-Oct-2002) Genome Sequencing Center,	4444 Forest Park
AUTHORS	TITLE	Parkway, St. Louis, MO 63108, USA	
COMMENT		On Oct 17, 2002 this sequence version replaced gi:23663310.	

```

----- Genome Center -----
Center: Washington University Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M BA0036C08

```

```
FEATURES
source
location/Qualifiers
1..247309
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
chromosome="1"
/clone="RP3-3C8"
BASE COUNT
72510 a 47853 c 48954 g 77992 t
```

Query Match	8.8%	Score 20;	DB 10;	length 247309;
Best Local Similarity	100.0%	Pred. No. 64;		
Matches	20;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
QY	106	TACCGATTTCTATTCTTG	125	
Db	236805	TACCGATTTCTATTCTTG	236824	

RESULT 18			
AB050528	AB050528	1717 bp	mRNA linear
LOCUS			PRI 27-OCT-2000
DEFINITION	Macaca fascicularis brain cDNA, clone:QMPA-19858.		
ACCESSION	AB050528		
VERSION	AB050528.1 GI:11041538		
KEYWORDS	fls (Full Insert sequence).		
SOURCE	Macaca fascicularis (crab-eating macaque)		
ORGANISM	Macaca fascicularis		

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (sites)
Osada,N., Hida,M., Kusuda,J., Tanuma,R., Ieki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished
2 (bases 1 to 1717)
Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
Direct Submission
Submitted (25-OCT-2000) Katsuyuki Hashimoto, National Institute
of Health Sciences, 1-1-1 Higashi, Tokyo 113, Japan

COMMENT

Crash: @crash@nih.gov, URL: http://www.nih.gov/yoken/genebank/
(E-mail: khash@nih.gov, IP: 207.46.130.100)
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181
Lab host: TOP10

```

FEATURES
    source
        1..1717
            location/Qualifiers
                R. Site1: DraIII (CAGCTGTTG)
                R. Site2: DraIII (CAGCATGTC)
                Description: 1st strand cDNA was primed with an oligo(dT) primer
                [ATGTCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesised
                using specific 5' and 3' primers and amplified by PCR. The PCR
                product was digested with SfiI and size selection was performed to
                exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
                into distinct DraIII sites of pM18S-F13. XhoI sites just outside
                the DraIII sites can be used to isolate the cDNA insert. Libraries
                were constructed by Sugano et al. (University of Tokyo, Institute of
                Medical Science). Custom primer used for sequencing
                ( 5' end primer [CTTCTGCTCTTAAAGCTGCC] ;
                3' end primer [CGACTCGAGCTCGACACA] ).

```

Query Match	8.3%	Score 19;	DB 9;	Length 1717;
Best Local Similarity	100.0%	Pred. No. 35;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

```

/mo1_type="mRNA"
/db_xref="taxon:9541"
/clone="Qmpa-19858"
/sex="male"
/tissue_type="brain parietal lobe"
/clone_lib="macaque brain cdna library Qmpa"
/dev_stage="adult"
BASE COUNT      505 a      402 g      489 t
ORIGIN
      321 c

```

	Query Match	Score	DB	Length
Best Local Similarity	100.0%	8.3%	19	1717
Matches	19	Pred. No.	35	
	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

RESULT 19	AE010268/c	10084 bp	DNA	linear	BCT 25-FEB-2001
LOCUS	AE010268	10084 bp	DNA	linear	BCT 25-FEB-2001
DEFINITION	Pyrococcus furiosus DSM 3658, section 143 of 173 of the complete genome.				
ACCESSION	AE010268	AE009950			
VERSION	AE010268.1	GI:18893862			
KEYWORDS					
SOURCE	Pyrococcus furiosus DSM 3658				

ORGANISM *Pyrococcus furiosus* DSM 3638
Archaea: Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.

REFERENCE
AUTHORS 1 (bases 1 to 10084)
TITLE Diruggiero, J., and Robb, F. T.
JOURNAL Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and
MEDLINE P. horikoshii inferred from complete genome sequences
PUBMED Genetics 152 (4), 1299-1305 (1999)
10430560

REFERENCE
AUTHORS 2 (bases 1 to 10084)
TITLE Robb, F. T., Maeder, D. L., Brown, J. R., Diruggiero, J., Stump, M. D.,
Yeh, R. K., Weis, R. B., and Dunn, D. M.
JOURNAL Genomic sequence of hyperthermophile, *Pyrococcus furiosus*:
MEDLINE implications for physiology and enzymology
PUBMED Meth. Enzymol. 330, 134-157 (2001)
21079003

REFERENCE
AUTHORS 3 (bases 1 to 10084)
TITLE Weiss, R. B., Dunn, D. M., Robb, F. T., and Brown, J. R.
JOURNAL The complete sequence of the *Pyrococcus furiosus* genome
AUTHORS 4 (bases 1 to 10084)
TITLE Weiss, R. B.
JOURNAL Direct Submission
JOURNAL Submitted (12-FEB-2002) Human Genetics, University of Utah, 20
South 2030 East, Salt Lake City, UT 84112, USA
FEATURES
source Location/Qualifiers
1. .10084
/organism="Pyrococcus furiosus DSM 3638"
/mol_type="genomic DNA"
/strain="DSM 3638"
/db_xref="taxon:186497"
68. .1078
/gene="PF1700"
68. .1078
/EC_number="4.6.1.4"
/note="Function Code: 1.2 Amino Acid Biosynthesis:
Aromatic amino acid family; (aroc)"
/codon_start=1
/transl_table=1
/product="chorismate synthase"
/protein_id="AAL81824.1"
/db_xref="GI:18893863"
/translation="MPGIRYNLELAELEERRKIGKISFTAKESDPIITISGVPRG
YTTGAPITVIVENKDVDSYEEIKDTPRPGHADYPAKIKYGFNDYGGGRFGRGLT
IGFVAGYFAKKILEPVGKIKAYIKRGVKEANMSIEEFNSPNYCPDEDFKRLM
LKEMESAKSGSDVGVVEVAIVNPGLGGVEDDIEADLASFPRIPAVKGVFGL
GPKYSEKRGSEVNDPVYIKNGEIQITNNHGGIIGITGMPIVARIAPKPTPIYLP
OKTYNLEBMEKETTITKLKGRPDSCIVPALPVEGAMAFVLADHFLRLAERYHLISK
EP"
1145. .1375
/gene="PF1701"
1145. .1375
/gene="PF1701"
/EC_number="5.4.99.5"
/note="Function Code: 1.2 Amino Acid Biosynthesis:
Aromatic amino acid family"
/codon_start=1
/transl_table=1
/product="chorismate mutase"
/protein_id="AAL81825.1"
/db_xref="GI:18893864"
/translation="MTLKLKREIDKIDNOIISLKKRLIEAQAIGIKIKELNPIE
DRKEEELVRAGFEIRFEKILEVSKDVQL"
1359. .2396
/gene="PF1702"
1359. .2396
/gene="PF1702"
/note="Function Code: 1.3 Amino Acid Biosynthesis:
Aspartate family; (asapB-2)"

gene
CDS
/codon_start=1
/transl_table=1
/product="aspartate aminotransferase"
/protein_id="AAL81826.1"
/db_xref="GI:18893865"
/translation="MKNVLELNKINEVKPEIRLDAGODPIPVANEIIEAINSIKRG
ETGVYSTTINEIREKIAIEVGSKEEIVGPGAKILLIAEIAMANKICVIAEYMAV
LILAKNPEREKIETTLNENSWPEINDVDLILNPNPTGKILPREKLELVE
VAEKIKILSDIEVAIEISFGFTPRELYENTVYKSGSKLVSMTGRLGVAIADKE
EIRKIPLESSTVCPVPRAGIKALERDELMKVSREYKRAELASKIRGLEF
YEPDGAFTFKPTIDGLEFVYKLERGSAFPGIAFGVYQNFIRISLSDLEKELN
IIEVAECESQ"
2381. .2827
/gene="PF1703"
2381. .2827
/gene="PF1703"
/EC_number="1.3.1.12"
/note="Function Code: 1.2 Amino Acid Biosynthesis:
Aromatic amino acid family"
/codon_start=1
/transl_table=1
/product="pyrophosphate dehydrogenase"
/protein_id="AAL81827.1"
/db_xref="GI:18893866"
/translation="MRTISGYGKMGKLFGRVLSSELNSEVRFYSHTLIDFSTLSEA
YEMSDVIMLSTIDNIKQIRLEIRLSINPNKIMFIDATPKKDLTEYQFPPEVK
VASVHMFPGGASFRGHLFYPCPCGQCKRRRRCKVYRIGSGS"
2832. .3155
/gene="PF1704"
2832. .3155
/gene="PF1704"
/note="Function Code: 16.4 Hypothetical"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAL81828.1"
/db_xref="GI:18893867"
/translation="MSVEEDRIVGVIGVPLGIAVYKLSTENNLERFGGTSPRFL
TIYKAVLNDGPRFIEVIRRSREIAEFLKSIDSVNVELSKVSRREIERAYEKFY
KILDP"
3155. .3890
/gene="PF1705"
3155. .3890
/gene="PF1705"
/EC_number="4.2.1.20"
/note="Function Code: 1.2 Amino Acid Biosynthesis:
Aromatic amino acid family"
/codon_start=1
/transl_table=1
/product="tryptophan synthetase alpha subunit"
/protein_id="AAL81829.1"
/db_xref="GI:18893868"
/translation="MFKDGLIPIYLTRAGDPDKOSTNLFLLADDEYAGAIELGTPSPDP
IADGKITQSHYRALKNGFKLEAFWIVVEFRHSSTPIVLMYTPITRAGRNPFLA
EAKASGVDDGLVVDLPVFAKKEPTEIAREGICTVPLAAPTDEBLKVYIDMTTFGV
YIVSLVGTGAREEIPKTAVIDLRAKRIICRNKVAAGVSGVSKREHVVSLKEGANGV
VGSALVKITIGEKGREATEPFLKKKVEELGI"
3890. .5049
/gene="PF1706"
3890. .5049
/gene="PF1706"
/EC_number="4.2.1.20"
/note="Function Code: 1.2 Amino Acid Biosynthesis:
Aromatic amino acid family; (trpB-2)"
/codon_start=1
/transl_table=1
/product="tryptophan synthase, subunit beta"
/protein_id="AAL81830.1"
/db_xref="GI:18893869"
/translation="MNFGRGCVVPEITLIEPLKELKAYKPKDDEEFNROLNYYLK
TWAGRPPLTYARLTKTGAKIYIKREDLVHGGAHKTNNAGQALLAFKMGKTRI
AETGAOGVATMAGALLGMKVDIWGAEDEVRQKNVFRMLLGNANVIVNSGSKT
LKDAINEALRDWATFEYTHYLGISVVGPHPTIVRFDOSVIGREAKQILBAEGOL

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10033
Center clone name: 360_N_22

* NOTE: This record contains 65 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 727: contig of 727 bp in length
* 728 827: gap of 100 bp
* 828 1483: contig of 656 bp in length
* 1484 1583: gap of 100 bp
* 1584 2318: contig of 735 bp in length
* 2319 2418: gap of 100 bp
* 2419 3147: contig of 729 bp in length
* 3148 3247: gap of 100 bp
* 3248 3962: contig of 715 bp in length
* 3963 4062: gap of 100 bp
* 4063 4763: contig of 701 bp in length
* 4764 4863: gap of 100 bp
* 4864 5560: contig of 697 bp in length
* 5561 6336: contig of 676 bp in length
* 6337 7143: contig of 100 bp
* 7144 7243: gap of 100 bp
* 7244 7927: contig of 684 bp in length
* 7928 8027: gap of 100 bp
* 8028 8768: contig of 741 bp in length
* 8769 8868: gap of 100 bp
* 8869 9597: contig of 729 bp in length
* 9598 9697: gap of 100 bp
* 9698 10412: contig of 715 bp in length
* 10413 10512: gap of 100 bp
* 10513 11165: contig of 653 bp in length
* 11166 11265: gap of 100 bp
* 11266 11959: contig of 694 bp in length
* 11960 12059: gap of 100 bp
* 12060 12765: contig of 706 bp in length
* 12766 12865: gap of 100 bp
* 12866 13557: contig of 692 bp in length
* 13558 13657: gap of 100 bp
* 13658 14357: contig of 700 bp in length
* 14358 14457: gap of 100 bp
* 14458 15186: contig of 729 bp in length
* 15187 15286: gap of 100 bp
* 15287 15994: contig of 708 bp in length
* 15995 16094: gap of 100 bp
* 16095 16787: contig of 693 bp in length
* 16788 16887: gap of 100 bp
* 16888 17662: gap of 100 bp
* 17663 18363: contig of 701 bp in length
* 18364 18463: gap of 100 bp
* 18464 19185: contig of 722 bp in length
* 19186 19285: gap of 100 bp
* 19286 19994: contig of 709 bp in length
* 19995 20094: gap of 100 bp
* 20095 20767: contig of 673 bp in length
* 20768 20867: gap of 100 bp
* 20868 21494: contig of 627 bp in length
* 21495 21594: gap of 100 bp
* 21595 22322: contig of 728 bp in length
* 22323 22422: gap of 100 bp
* 22423 23153: contig of 731 bp in length

FEATURES

23154 23253: gap of 100 bp
23254 23981: contig of 728 bp in length
23982 24081: gap of 100 bp
24082 24792: contig of 711 bp in length
24793 24892: gap of 100 bp
24893 25538: contig of 646 bp in length
25539 25638: gap of 100 bp
25639 26376: contig of 737 bp in length
26376 26475: gap of 100 bp
26475 27184: contig of 709 bp in length
27185 27284: gap of 100 bp
27285 28019: contig of 735 bp in length
28020 28119: gap of 100 bp
28120 28816: contig of 697 bp in length
28817 28916: gap of 100 bp
28917 29629: contig of 713 bp in length
29630 29729: gap of 100 bp
29730 30404: contig of 675 bp in length
30405 30504: gap of 100 bp
30505 31200: contig of 696 bp in length
31201 31300: gap of 100 bp
31301 31971: contig of 671 bp in length
31972 32071: gap of 100 bp
32072 32777: contig of 706 bp in length
32778 32877: gap of 100 bp
32878 33617: contig of 740 bp in length
33618 33717: gap of 100 bp
33718 34439: contig of 722 bp in length
34440 34539: gap of 100 bp
34540 35230: contig of 691 bp in length
35231 35330: gap of 100 bp
35331 36046: contig of 716 bp in length
36047 36146: gap of 100 bp
36147 36819: contig of 673 bp in length
36820 36919: gap of 100 bp
36920 37622: contig of 703 bp in length
37623 37722: gap of 100 bp
37723 38425: contig of 703 bp in length
38426 38525: gap of 100 bp
38526 39144: contig of 619 bp in length
39145 39244: gap of 100 bp
39245 39975: contig of 731 bp in length
39976 40075: gap of 100 bp
40076 40774: contig of 699 bp in length
40874 40874: gap of 100 bp
40875 41542: contig of 668 bp in length
41543 41642: gap of 100 bp
41643 42358: contig of 716 bp in length
42359 42458: gap of 100 bp
42459 43150: contig of 692 bp in length
43151 43250: gap of 100 bp
43251 43954: contig of 704 bp in length
43955 44054: gap of 100 bp
44055 44711: contig of 657 bp in length
44712 44811: gap of 100 bp
44812 45501: contig of 690 bp in length
45502 45601: gap of 100 bp
45602 46328: contig of 727 bp in length
46329 46428: gap of 100 bp
46429 47151: contig of 723 bp in length
47152 47251: gap of 100 bp
47252 47969: contig of 718 bp in length
47970 48069: gap of 100 bp
48070 48775: contig of 706 bp in length
48776 48875: gap of 100 bp
48876 49516: contig of 641 bp in length
49517 49616: gap of 100 bp
49617 50304: contig of 688 bp in length
50305 50404: gap of 100 bp
50405 51108: contig of 704 bp in length
51109 51208: gap of 100 bp
51209 51896: contig of 688 bp in length
Location/Qualifiers

source

1. 51896
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-360N22"
/clone_11b="RP11-360N22" Human Male BAC"

Query Match 8.3%; Score 19; DB 2; Length 51896;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 ATTTCTATTCTTGCCCTC 129
|||||
Db 6901 ATTTCTATTCTTGCCCTC 6919

RESULT 22
AL161614 70946 bp DNA linear PRI 14-NOV-2001
LOCUS Human DNA sequence from clone Rp11-346L13 on chromosome
DEFINITION 13q14.3-21.31, complete sequence.
ACCESSION AL161614 GI:16944883
VERSION AL161614.16
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 70946)
Martin, S.
Submitted (14-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 15, 2001 this sequence version replaced gi:16304689.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rp11-346L13 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-346L13. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone Rp11-471M10 is at 68947 in this
sequence. The true right end of clone Rp11-181D10 is at 2000 in
this sequence.

FEATURES
Source
1..70946
/organism="Homo sapiens"
/mol_type="genomic DNA"

/db_xref="taxon:9606"
/chromosome="13"
/map="q14.3-21.31"
/clone="RP11-346L13"
/clone_11b="RP11-11.2"
3097.3326
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
16413.16577
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
3114.31383
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."

BASE COUNT 20505 a 15606 c 15147 g 19688 t
ORIGIN

Query Match 8.3%; Score 19; DB 9; Length 70946;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 GATTTCATTCTTGCCCT 128
|||||
Db 19979 GATTTCATTCTTGCCCT 19961

RESULT 23
AC068522 146952 bp DNA linear PRI 30-MAY-2002
LOCUS Homo sapiens chromosome 8, clone Rp11-23K11, complete sequence.
DEFINITION AC068522
ACCESSION AC068522.7 GI:21283337
VERSION AC068522.7
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 146952)
Birren, B., Linton, L., Nussbaum, C., and Lander, E.
Homo sapiens chromosome 8, clone Rp11-23K11
Unpublished
2 (bases 1 to 146952)
Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karataa, A.,
Klein, J., Larocque, K., Lamazeres, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., Medwan, P., McGurk, A., McKernan, K., McPherson, R.,
Meldrum, J., Menous, J., Mihova, T., Miranda, C., Mieng, V., Morrow, J.,
Murphy, T., Naylor, L., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,
Teisfey, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, K., Wyman, D., Ye, W. J.,
Young, G., Zaitoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 146952)
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kats, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McKernan, K., Meldrum, J., Menus, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicot, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, J., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strause, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 146952)

Birtten, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barina, N., Baetien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McKernan, K., Meldrum, J., Menus, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicot, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, J., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strause, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (30-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 30, 2002 this sequence version replaced g1:20331012.
All repeats were identified using RepeatMasker:
http://fpc.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L7815
Center Clone name: 23_K_11

FEATURES

SOURCE

Location/Qualifiers

1. 146952

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-23K11"

/clone_lib="RP11-11 Human Male BAC"

/repeat_region
complement(255..413)

/rpt_family="MIR"

repeat_region
417..474

/rpt_family="MLT1F"
557..742

/rpt_family="MLT1F"

repeat_region
771..794

/rpt_family="(TTG)n"

repeat_region
2054..2077

/rpt_family="(TG)n"

complement(2397..2695)

/rpt_family="AluSc"

repeat_region
2917..3059

/rpt_family="FLAM_C"

repeat_region
3199..3319

/rpt_family="L2"

repeat_region
complement(3685..4440)

/rpt_family="L1MA9"

repeat_region
4556..4596

/rpt_family="(CA)n"

repeat_region
4747..5070

/rpt_family="MLT1B"

repeat_region
5071..6326

/rpt_family="L1HS"

repeat_region
6346..6419

/rpt_family="MLT1B"

repeat_region
complement(8828..9137)

/rpt_family="AluSx"

repeat_region
9140..9170

/rpt_family="AT_rich"

repeat_region
9315..9433

/rpt_family="MIR3"

repeat_region
9946..10153

/rpt_family="MIR"

repeat_region
complement(10341..10664)

/rpt_family="AluSx"

repeat_region
complement(10803..11140)

/rpt_family="MERS8B"

repeat_region
11542..11853

/rpt_family="AluSx"

repeat_region
11950..11972

/rpt_family="AT_rich"

repeat_region
complement(12690..12813)

/rpt_family="FLAM_A"

repeat_region
14222..14272

/rpt_family="GA-rich"

repeat_region
complement(14439..14750)

/rpt_family="AluJb"

repeat_region
complement(15243..15341)

/rpt_family="MIR"

repeat_region
15398..15706

/rpt_family="AluSx"

repeat_region
15977..16133

/rpt_family="MIR"

repeat_region
16534..16844

/rpt_family="AluSg"

repeat_region
complement(16884..17207)

/rpt_family="AluY"

repeat_region
17956..18150

/rpt_family="MIR"

repeat_region
complement(19195..19360)

/rpt_family="FRAM"

repeat_region
complement(20389..20507)

/rpt_family="L2"

repeat_region
complement(22803..23068)

/rpt_family="L2"

repeat_region
complement(23151..23306)

/rpt_family="L1ME3A"

repeat_region
23485..23519

/rpt_family="AT_rich"

repeat_region
23687..24028

/rpt_family="THE1B"

repeat_region
24029..24453

/rpt_family="THE1B-int"

complement(24454..24760)

/rpt_family="AluSg"

repeat_region 24761..25082
/rpt_family="THE1B-int"
repeat_region 25107..25128
/rpt_family="(T)n"
repeat_region 26515..26614
/rpt_family="CT-rich"
repeat_region 26619..26646
/rpt_family="(CA)n"
repeat_region complement(26647..26752)
/rpt_family="L2"
repeat_region 27726..27924
/rpt_family="MIR"

Query Match 8.3%; Score 19; DB 9; Length 146952;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 ATTTCTATTCTTGCTC 129
|||||
Db 138439 ATTTCTATTCTTGCTC 138421

RESULT 24
HSJ520B18 148385 bp DNA linear PRI 31-JUL-2000
LOCUS
DEFINITION Human DNA sequence from clone RP3-520B18 on chromosome 6p24.1-25.3
Contains the 5' end of PARS1 (phenylalanine-tRNA synthetase), a
pseudogene similar to microtubule-associated protein 1A/1B light
chain 3, a CpG island, ESTs, STS and GSSs, complete sequence.
ALI21978
ALI21978.5 GI:8247628
HTG: CpG island; tRNA synthetase.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 148385)
Dunn, M.
Direct Submission
Submitted (27-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonequery@sanger.ac.uk
On Jun 4, 2000 this sequence version replaced gi:8247028.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP3-520B18 is from the library RPCI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-520B18.
FEATURES
Source
1..148385
/organism="Homo sapiens"
/mol_type="genomic DNA"

/db_xref="taxon:9606"
/chromosome="6"
/map="p24.1-25.3"
/clone="RP3-520B18"
/clone_1db="RPCI-3"
49..158
/note="L1B repeat: matches 5690..5801 of consensus"
303..474
/note="L1B repeat: matches 5352..5526 of consensus"
873..1159
/note="AluSp repeat: matches 1..295 of consensus"
1562..1884
/note="L2 repeat: matches 1475..1802 of consensus"
1980..2628
/note="L1P8 repeat: matches 5514..6163 of consensus"
2764..3060
/note="AluX repeat: matches 1..296 of consensus"
3202..3512
/note="AluY repeat: matches 1..307 of consensus"
5478..5708
/note="AluSg repeat: matches 1..302 of consensus"
8139..8257
/note="FLM_A repeat: matches 1..121 of consensus"
8904..9405
/note="match: GSS: Em:AQ713403"
8904..9219
/note="match: GSS: Em:B43761"
8911..9405
/note="match: GSS: Em:AQ057163"
9636..9808
/note="AluSg/x repeat: matches 123..230 of consensus"
10666..10968
/note="AluX repeat: matches 1..303 of consensus"
11149..11444
/note="Alu repeat: matches 1..287 of consensus"
11612..12104
/note="MER1A repeat: matches 1..527 of consensus"
14547..14889
/note="match: STS: Em:HS163H16S"
15496..15673
/note="WIR repeat: matches 61..248 of consensus"
15966..15816
/note="WIR repeat: matches 131..256 of consensus"
16119..16241
/note="L2 repeat: matches 2629..2750 of consensus"
16558..16854
/note="AluY repeat: matches 1..297 of consensus"
16921..17206
/note="AluSg repeat: matches 1..287 of consensus"
17250..17555
/note="AluY8 repeat: matches 1..306 of consensus"
17607..18098
/note="match: GSS: Em:AQ498962"
18039..18523
/note="match: GSS: Em:AQ143295"
18524..18874
/note="MER47A repeat: matches 2..366 of consensus"
19573..19646
/note="37 copies 2 mer at 83% conserved"
19884..19966
/note="L2 repeat: matches 2624..2712 of consensus"
20286..20497
/note="L2 repeat: matches 2178..2427 of consensus"
20794..21288
/note="match: GSS: Em:AQ0808524"
20850..20964
/note="MIR repeat: matches 15..134 of consensus"
complement(21176..21717)
/note="match: GSS: Em:B95444"
21281..21473
/note="L2 repeat: matches 1545..1741 of consensus"
complement(21298..21740)
/note="match: GSS: Em:AQ129989"

```
repeat_region 22649..23017 /note="L2 repeat: matches 1900..2294 of consensus"
repeat_region 23092..23404 /note="L2 repeat: matches 2405..2705 of consensus"
repeat_region 24603..24716 /note="57 copies 2 mer tt 59% conserved"
repeat_region 25819..26224 /note="MSTR repeat: matches 1..426 of consensus"
repeat_region 26760..26845 /note="MIR repeat: matches 25..122 of consensus"
repeat_region 26913..27042 /note="L2 repeat: matches 2299..2432 of consensus"
repeat_region 27793..29016 /note="HSMAR1 repeat: matches 1..1285 of consensus"
repeat_region 29670..29940 /note="Char1:elb repeat: matches 34..334 of consensus"
repeat_region 29995..30093 /note="Char1:elb repeat: matches 421..521 of consensus"
repeat_region 30195..30342 /note="MER20 repeat: matches 42..217 of consensus"
repeat_region 30437..30498 /note="MIR repeat: matches 189..250 of consensus"
repeat_region 30516..30580 /note="MER20 repeat: matches 154..217 of consensus"
repeat_region 30691..31003 /note="ALUSq repeat: matches 1..313 of consensus"
repeat_region 31018..31057 /note="Char1:el1 repeat: matches 2720..2759 of consensus"
repeat_region 31059..31149 /note="MIR repeat: matches 70..168 of consensus"
repeat_region 31159..31815 /note="L2 repeat: matches 1688..2371 of consensus"
repeat_region 32006..32327 /note="MER1B repeat: matches 1..337 of consensus"
repeat_region 32415..32649 /note="L2 repeat: matches 2291..2532 of consensus"
repeat_region 32822..32998 /note="MIR repeat: matches 76..262 of consensus"
repeat_region 34981..36185 /note="Cpg island"
repeat_region 36754..37061 /note="ALUSx repeat: matches 1..307 of consensus"
repeat_region 38414..38743 /note="match: GSS: Em:AQ133010"
repeat_region 38974..39189 /note="ALUdb repeat: matches 2..294 of consensus"
repeat_region 39288..39581 /note="ALUSx repeat: matches 1..303 of consensus"
repeat_region 40838..40868 /note="ALUdb repeat: matches 1..293 of consensus"
repeat_region 40869..41011 /note="MIR repeat: matches 122..151 of consensus"
repeat_region 41012..41060 /note="FLAN C repeat: matches 3..133 of consensus"
repeat_region 41012..41060 /note="MIR repeat: matches 69..122 of consensus"
repeat_region 41952..42253 /note="match: GSS: Em:AQ142457"
repeat_region 42254..43416 /note="ALU repeat: matches 1..300 of consensus"
repeat_region 43417..43944 /note="L1P4S repeat: matches 4420..5588 of consensus"
repeat_region 43946..43987 /note="L1P4S repeat: matches 5585..6143 of consensus"
repeat_region 43946..43987 /note="21 copies 2 mer aa 81% conserved"
misc_feature complement(44030..44549)
misc_feature /note="match: STS: Em:G59196
match: GSS: Em:AQ585941"
misc_feature complement(44282..44623)
misc_feature /note="match: GSS: Em:AQ776653"
repeat_region 45282..45439
```

```
repeat_region /note="L2 repeat: matches 2328..2492 of consensus"
repeat_region 46044..46305 /note="ALUSg repeat: matches 39..300 of consensus"
repeat_region 46439..46658 /note="MIR44A repeat: matches 478..326 of consensus"

Query Match 8.3%; Score 19; DB 9; Length 148185;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 146 TCACCTGGATTTCCTCA 164
Db 4442 TCACCTGGATTTCCTCA 4460

RESULT 25
AC093533/c 154814 bp DNA linear PRI 26-MAR-2003
DEFINITION Homo sapiens chromosome 5 clone RP11-46C20, complete sequence.
ACCESSION AC093533 AC011082
VERSION AC093533.2 GI:29251572
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 154814)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 154814)
DOE Joint Genome Institute.
Direct Submission
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 154814)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (26-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On or before Mar 26, 2003 this sequence version replaced
GI:6006258, GI:15383822.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Fishing completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated total Number of Errors is 0.
Location/Qualifiers
1..154814
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-46C20"

BASE COUNT 51008 a 26529 c 26123 g 51154 t
ORIGIN

Query Match 8.3%; Score 19; DB 9; Length 154814;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 ATATTGCTCCACCTTC 224
Db 111611 ATATTGCTCCACCTTC 111593

RESULT 26
AC016174/c 161304 bp DNA linear HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-20K23, WORKING DRAFT SEQUENCE, 19 unordered
pieces.
ACCESSION AC016174
```

VERSION AC016174.4 GI:7249016
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 161304)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-20K23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 161304)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dekrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J.,
Galagan,J.C., Gardner,S., Grant,G., Hages,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kahn,L., Karatsis,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McBwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,D.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Sudrmanian,A., Talamas,U.,
Tsifaye,S., Turrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wymann,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6721350.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13921
Center clone name: 20_K.23
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 14819 bases at least Q40
Consensus quality: 154100 bases at least Q30
Consensus quality: 156444 bases at least Q20
Insert size: 15900; agarose-fp
Insert size: 15904; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1227: contig of 1227 bp in length
* 1228 1327: gap of 100 bp
* 1328 3446: contig of 219 bp in length
* 3447 3547: gap of 100 bp
* 3547 5141: contig of 1595 bp in length
* 5142 5242: gap of 100 bp
* 5242 7279: contig of 2038 bp in length
* 7280 7379: gap of 100 bp
* 7380 9749: contig of 2370 bp in length
* 9750 9849: gap of 100 bp
* 9850 12547: contig of 2698 bp in length
* 12548 12647: gap of 100 bp
* 12648 15286: contig of 2639 bp in length

FEATURES
source
1. 15287 15386: gap of 100 bp
* 15387 17760: contig of 2374 bp in length
* 17761 17860: gap of 100 bp
* 17861 19830: contig of 1970 bp in length
* 19831 19931: gap of 100 bp
* 19931 23207: contig of 3277 bp in length
* 23208 23307: gap of 100 bp
* 23308 29134: contig of 5827 bp in length
* 29135 29235: gap of 100 bp
* 29235 34937: contig of 5703 bp in length
* 34938 35037: gap of 100 bp
* 35038 40907: contig of 5870 bp in length
* 40908 41007: gap of 100 bp
* 41008 51635: contig of 10628 bp in length
* 51636 51735: gap of 100 bp
* 51736 63754: contig of 11919 bp in length
* 63755 78005: gap of 100 bp
* 78006 78105: gap of 100 bp
* 78106 99941: contig of 21836 bp in length
* 99942 100041: gap of 100 bp
* 100042 129873: contig of 29832 bp in length
* 129874 129973: gap of 100 bp
* 129974 161304: contig of 31331 bp in length.
Location/Qualifiers
1. 161304
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-20K23"
/clone_lib="PFCI-11 Human Male BAC"
1. 1227
/note="assembly_fragment"
1328. 3446
/note="assembly_fragment"
3547. 5141
/note="assembly_fragment"
5242. 7279
/note="assembly_fragment"
clone_end:T7
vector_side:left"
7380. 9749
/note="assembly_fragment"
9850. 12547
/note="assembly_fragment"
12548. 15286
/note="assembly_fragment"
15387. 17760
/note="assembly_fragment"
17861. 19830
/note="assembly_fragment"
19931. 23207
/note="assembly_fragment"
23308. 29134
/note="assembly_fragment"
29235. 34937
/note="assembly_fragment"
35038. 40907
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
41008. 51635
/note="assembly_fragment"
51736. 63754
/note="assembly_fragment"
63755. 78005
/note="assembly_fragment"
78106. 99941
/note="assembly_fragment"
100042. 129873
/note="assembly_fragment"
129974. 161304
/note="assembly_fragment"

BASE COUNT 52553 a 28110 c 28225 g 50612 t 1804 others
ORIGIN

Query Match 8.3%; Score 19; DB 2; Length 161304;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ATTTCATCTCTGCGCTC 129
|||||
Db 53189 ATTTCATCTCTGCGCTC 53171

RESULT 27
AL954849/c 162939 bp DNA linear VRT 25-APR-2003
LOCUS Zebrafish DNA sequence from clone DKEX-7N10 in linkage group 7,
DEFINITION complete sequence.
ACCESSION AL954849 GI:30141725
VERSION AL954849
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 162939)
Coryby, N.
REFERENCE Direct Submission
Submitted (25-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Apr 26, 2003 this sequence version replaced gi:30024500.
COMMENT -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhong Bao and Sean Eddy, submitted), and those beginning 'dtr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEX-7N10 is from a Zebrafish BAC library

FEATURES
Source Location/Qualifiers

1..162939
/organism="Danio rerio"
/mol_type="Genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-7N10"
/clone_id="DanioKey"
BASE COUNT 53983 a 27956 c 27092 g 53908 t
ORIGIN

Query Match 8.3%; Score 19; DB 5; Length 162939;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 TTTCCTGCTCGTATTGTC 137
|||||
Db 127994 TTTCCTGCTCGTATTGTC 127976

RESULT 28
AC108034 164396 bp DNA linear PRI 23-FEB-2002
LOCUS Homo sapiens BAC clone RP11-137016 from 4, complete sequence.
DEFINITION AC108034
ACCESSION AC108034 GI:18693544
VERSION AC108034.2
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164396)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE 9847074
PUBMED
REFERENCE 2 (bases 1 to 164396)
Levy, A. and Kozlowski, A.
The sequence of Homo sapiens BAC clone RP11-137016
Unpublished (2001)
3 (bases 1 to 164396)
REFERENCE
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 164396)
Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 164396)
Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 16, 2002 this sequence version replaced gi:18308807.
COMMENT -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0137J16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frenken, E.,
Tateno, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. *Genomics* 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-137J16;
Actual end is at base position 164396 of RP11-137J16.

Location/Qualifiers

```
FEATURES
SOURCE
1.164396
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-137J16"
/clone_lib="RPCI-11"
1..98
repeat_region
/rpt_family="MALR"
494..541
repeat_region
/rpt_family="A-rich"
211..2469
repeat_region
/rpt_family="MALR"
3376..3720
repeat_region
/rpt_family="MALR"
3736..3957
repeat_region
/rpt_family="ERV1"
3961..4441
repeat_region
/rpt_family="ERV1"
4442..4996
repeat_region
/rpt_family="ERV1"
4999..5117
repeat_region
/rpt_family="MER1_type"
5119..5473
repeat_region
/rpt_family="MER1_type"
5474..5854
repeat_region
/rpt_family="MALR"
5855..6303
repeat_region
/rpt_family="MER1_type"
6304..7348
repeat_region
/rpt_family="MER2_type"
7349..7770
repeat_region
/rpt_family="L1"
7771..8071
repeat_region
/rpt_family="MER2_type"
8072..8288
repeat_region
/rpt_family="A1u"
8306..8432
repeat_region
/rpt_family="MER2_type"
8712..8966
repeat_region
/rpt_family="L2"
9854..10129
repeat_region
/rpt_family="MALR"
13738..13770
repeat_region
/rpt_family="A1u"
13995..14472
repeat_region
/rpt_family="MALR"
14649..14675
repeat_region
/rpt_family="CA)n"
14831..14937
repeat_region
/rpt_family="MIR"
14938..15236
repeat_region
/rpt_family="A1u"
15237..15318
repeat_region
```

```
/rpt_family="MIR"
15370..15449
repeat_region
/rpt_family="MALR"
16344..16651
repeat_region
/rpt_family="A1u"
16703..16745
repeat_region
/rpt_family="AT-rich"
17302..17334
repeat_region
/rpt_family="AT-rich"
18160..18199
repeat_region
/rpt_family="CA)n"
18315..18407
repeat_region
/rpt_family="MIR"
18556..18805
repeat_region
/rpt_family="ERV1"
18806..19139
repeat_region
/rpt_family="MALR"
19351..19495
repeat_region
/rpt_family="ERV1"
19496..19853
repeat_region
/rpt_family="MALR"
21433..21764
repeat_region
/rpt_family="MALR"
21785..21936
repeat_region
/rpt_family="ERV1"
21937..22034
repeat_region
/rpt_family="ERV1"
22352..22374
repeat_region
/rpt_family="AT-rich"
22376..22637
repeat_region
/rpt_family="L1"
22638..22667
repeat_region
/rpt_family="TA)n"
22668..23500
repeat_region
/rpt_family="L1"
23501..23806
repeat_region
/rpt_family="A1u"
23807..23957
repeat_region
/rpt_family="L1"
25504..25918
repeat_region
/rpt_family="L2"
25991..26291
repeat_region
/rpt_family="A1u"
26443..26669
repeat_region
/rpt_family="MIR"
26866..27577
repeat_region
/rpt_family="L1"
27745..29220
repeat_region
/rpt_family="L1"
29205..30237
repeat_region
/rpt_family="L1"
30238..30265
repeat_region
/rpt_family="TA)n"
30266..30393
repeat_region
/rpt_family="L1"
30395..30939
repeat_region
/rpt_family="ERV1"
30945..31238
repeat_region
/rpt_family="A1u"
31239..31287
repeat_region
/rpt_family="AT-rich"
31331..31500
repeat_region
/rpt_family="L1"
31501..31789
repeat_region
/rpt_family="A1u"
31790..32506
repeat_region
```

Query Match 8 3%; Score 19; DB 9; Length 164396;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Gaps 0;
0Y 111 ATTTCTATTCTGACCTC 129
|||||

DB 35568 ATTTCTATTCTTCGCTC 35586

RESULT 29
AC116849
LOCUS
DEFINITION Mus musculus clone RP24-395F1, WORKING DRAFT SEQUENCE, 13 unordered pieces.

AC116849 176785 bp DNA linear HTG 22-FEB-2003
AC116849 GI:28467354
HTG, HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 176785)
JOURNAL Mus musculus, clone RP24-395F1
REFERENCE
AUTHORS 2 (bases 1 to 176785)
TITLE Unpublished

REFERENCE
AUTHORS Anderson, S., Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarty, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS 3 (bases 1 to 176785)
TITLE Unpublished

REFERENCE
AUTHORS Birren, B., Nussbaum, C., Lander, E., Aboueleil, A., Allen, N., Anderson, S., Archchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J., Choquel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearrellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarty, M., Meldrum, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Neuyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rieback, M., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 22, 2003 this sequence version replaced g1:20800051. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L25615
Project clone name: 395_F_1

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173096 bases at least Q40
Consensus quality: 174668 bases at least Q30
Consensus quality: 175223 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 175585; sum-of-coverage
Quality coverage: 8.9 in Q20 bases; agarose-fp
Quality coverage: 8.5 in Q20 bases; sum-of-coverage

----- NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 88954: contig of 88954 bp in length
* 88955 88954: gap of 100 bp
* 89055 89774: contig of 720 bp in length
* 89775 89774: gap of 100 bp
* 89875 90628: contig of 754 bp in length
* 90629 90728: gap of 100 bp
* 90729 91644: contig of 916 bp in length
* 91645 91745: gap of 100 bp
* 91745 92427: contig of 683 bp in length
* 92428 92527: gap of 100 bp
* 92528 93177: contig of 650 bp in length
* 93178 93277: gap of 100 bp
* 93278 93908: contig of 631 bp in length
* 93909 94008: gap of 100 bp
* 94009 94751: contig of 743 bp in length
* 94752 94851: gap of 100 bp
* 94852 96573: contig of 1722 bp in length
* 96574 96673: gap of 100 bp
* 96674 98830: contig of 2157 bp in length
* 98831 98930: gap of 100 bp
* 98931 116439: contig of 17509 bp in length
* 116440 116539: gap of 100 bp
* 116540 154762: contig of 38223 bp in length
* 154763 154862: gap of 100 bp
* 154863 176785: contig of 21923 bp in length.

FEATURES

source
1. 176785
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-395F1"
/clone_id="RP24-395F1"
1. 88954
/note="assembly_fragment
clone end:SP6
vector_start:left"
misc_feature
89055..89774
/note="assembly_fragment"
misc_feature
89875..90628
/note="assembly_fragment"
misc_feature
90729..91644
/note="assembly_fragment"
misc_feature
91745..92427
/note="assembly_fragment"
misc_feature
92528..93177

```

misc_feature      /note="assembly_fragment"  
93278. .93908  
/note="assembly_fragment"  
misc_feature      94009. .94751  
/note="assembly_fragment"  
misc_feature      94852. .96573  
/note="assembly_fragment"  
misc_feature      96674. .98830  
/note="assembly_fragment"  
misc_feature      98931. .116439  
/note="assembly_fragment"  
misc_feature      116540. .154762  
/note="assembly_fragment"  
misc_feature      154863. .176785  
/note="assembly_fragment"  
clone_end:r7  
vector_side:right"  
BASE COUNT       47009 a 39536 c 39976 g 49062 t 1202 others  
ORIGIN  
Query Match      8.3%; Score 19; DB 2; Length 176785;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBMC
Center clone name: RP11-466116
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye 3 of reads
Chemistry: Dye-terminator Big Dye 3 of reads
Assembly program: Phred; version 0.990329
Consensus quality: 171549 bases at least Q40
Consensus quality: 181986 bases at least Q30
Consensus quality: 188593 bases at least Q20
Estimated insert size: 208123; sum-of-coverage estimation
Quality coverage: 3x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2654: contig of 2654 bp in length
2655
2754: gap of unknown length
2755
5039: contig of 2285 bp in length
5040
5139: gap of unknown length
5140
7837: contig of 2698 bp in length
7838
7937: gap of unknown length
7938
10127: contig of 2190 bp in length
10128
10227: gap of unknown length
10228
12612: contig of 2385 bp in length
12613
12712: gap of unknown length
12713
15508: contig of 2796 bp in length
15509
15608: gap of unknown length
15609
18028: contig of 2420 bp in length
18029
18128: gap of unknown length
18129
20134: contig of 2006 bp in length
20135
20234: gap of unknown length
20235
22632: contig of 2398 bp in length
22633
22732: gap of unknown length
22733
26184: contig of 3452 bp in length
26185
26284: gap of unknown length
26285
29499: contig of 3215 bp in length
29500
29599: gap of unknown length
32924: contig of 3325 bp in length
32925
33024: gap of unknown length
33025
36645: contig of 3621 bp in length
36646
36745: gap of unknown length
40074: contig of 3329 bp in length
40075
40174: gap of unknown length
40175
46122: contig of 5948 bp in length
46123
46222: gap of unknown length
51842: contig of 5620 bp in length
51843
51942: gap of unknown length
58190: contig of 6248 bp in length
58191
58290: gap of unknown length
58291
66300: contig of 8010 bp in length
66301
66400: gap of unknown length
66401
75501: contig of 9101 bp in length
75502
85178: contig of 9577 bp in length
85179
85278: gap of unknown length
85279
91453: contig of 6175 bp in length
91454
91553: gap of unknown length
91554
101474: contig of 9921 bp in length
101475
101574: gap of unknown length
114100: contig of 12526 bp in length

```

```

* 114101 114200: gap of unknown length
* 114201 128916: contig of 14716 bp in length
* 128917 129016: gap of unknown length
* 129017 146607: contig of 17591 bp in length
* 146608 146707: gap of unknown length
* 146708 166453: contig of 19646 bp in length
* 166454 166453: gap of unknown length
* 166454 192409: contig of 25956 bp in length.
  Location/Qualifiers
    1..192409
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="3"
      /clone="RP11-466116"
BASE COUNT 57427 a 38727 c 38414 g 55163 t 2678 others
ORIGIN
Query Match 8.3%; Score 19; DB 2; Length 192409;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 111 ATTTCTATTCTTGCTC 129
Db 60081 ATTTCTATTCTTGCTC 60099
-----
RESULT 32
AC092700/c
LOCUS 20186 bp DNA linear PRI 06-JUN-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-661A3, complete sequence.
ACCESSION AC092700
VERSION AC092700.2 GI:21327455
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 20186)
Homo sapiens chromosome 8, clone RP11-661A3
Unpublished
2 (bases 1 to 20186)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Brown,A., Camarata,J., Campiano,A., Chang,S., Charao,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dairellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Harford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamasares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marguis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meidlin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Polata,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,B., Schupack,R.,
Seaman,S., Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strause,N., Subramanian,A., Talamas,J., Testaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zaitoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
Research, 330 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 20186)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Charao,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,

```


Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardy, A., Girdler, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatae, A., Kells, C., Larroque, K.,
 Lamazeres, R., Landers, T., Lenochky, J., Levine, R., Lindblad-Ton, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, D.,
 Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, O., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 6, 2002 this sequence version replaced gi:149711446.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seg.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L11751

Center clone name: 661_A_3

FEATURES

source

```

-----
Location/Qualifiers
1. .20186
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-661A3"
/clone_id="RP11-11 Human Male BAC"
4356. .4523
/rpt_family="HALb"
complement(4625. .4906)
/rpt_family="AluIo"
8941. .9014
/rpt_family="AluJ/FRAM"
9122. .9238
/rpt_family=" (TA)n"
11059. .11102
/rpt_family="L3"
complement(11461. .11765)
/rpt_family="AluSx"
12218. .12291
/rpt_family="cRNA-Glu-GAG"
12584. .12885
/rpt_family="AluSx"
16513. .16567
/rpt_family="L2"
16626. .16926
/rpt_family="AluY"
18349. .18494
/rpt_family="AluSg/x"
18705. .19009
/rpt_family="AluSx"
complement(20040. .20229)
/rpt_family="MIR"
20313. .20612
/rpt_family="AluSc"
20813. .21145
/rpt_family="L1MB4A"
repeat_region
/rpt_family="AT_rich"
22622. .22652
repeat_region
/rpt_family="L1MB4A"
22708. .22812
repeat_region
/rpt_family="AT_rich"
24360. .24382
repeat_region
/rpt_family=" (TA)n"
24482. .24545
repeat_region
complement(24715. .25019)
/rpt_family="AluDb"
25673. .25879
/rpt_family="AluSg/x"
26165. .26219
/rpt_family=" (TA)n"
complement(26405. .26553)
/rpt_family="AluDb"
26671. .26899
/rpt_family="MIR"
27947. .28028
/rpt_family="MIR"
30867. .31239
/rpt_family="L1MCC"
31240. .31487
/rpt_family="AluDb"
31488. .31862
/rpt_family="L1MCC"
31923. .32003
/rpt_family="MER21B"
32024. .32771
/rpt_family="MER21B"
32826. .33568
/rpt_family="L1MCC"
33580. .33721
/rpt_family="MER7A"
complement(33722. .33950)
/rpt_family="MER46A"
33951. .34109
/rpt_family="MER7A"
34110. .34303
/rpt_family="LTR4"
34483. .34764
/rpt_family="LTR4"
34829. .36974
/rpt_family="L1MCC"
37129. .38284
/rpt_family="L1PA6"
38328. .38382
/rpt_family="AT_rich"
38383. .38550
/rpt_family="L1MCC"
38555. .38600
/rpt_family=" (CA)n"
38633. .38743
/rpt_family="L1MCC"
39499. .39768
/rpt_family="L1MCC"
39772. .39960
/rpt_family="L1MCC"
complement(40150. .40364)
/rpt_family="L2"
40482. .40510
/rpt_family="AT_rich"
complement(40913. .42148)
/rpt_family="L1PA15"
42223. .42374
/rpt_family="AluDb"
complement(42375. .42667)
/rpt_family="AluSg"
complement(42668. .42807)
/rpt_family="AluDb"
complement(43919. .44010)
/rpt_family="L1"
complement(44011. .44322)
repeat_region

```

```
repeat_region /rpt family="AlusX"
complement(44323..44921)
repeat_region /rpt family="L1"
44923..45311
repeat_region /rpt family="MLT2B1"
45342..45403
repeat_region /rpt family="(TATG)n"
45405..45463
repeat_region /rpt family="MLT2D"
complement(45464..45891)

Query Match 8.3%; Score 19; DB 9; Length 201866;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ATTTCTATTCTTGCTC 129
|||||
Db 36734 ATTTCTATTCTTGCTC 36716

RESULT 33
AC024249/c 202983 bp DNA linear HTG 07-JUN-2000
LOCUS Homo sapiens chromosome 8 clone RP11-661A3, WORKING DRAFT SEQUENCE,
DEFINITION 40 unordered pieces.
ACCESSION AC024249 GI:7631100
VERSION HTG; HTG_PHASE1; HTG_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 202983)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 202983)
Waterston, R.H.
Direct Submission
Submitted (28-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 21, 2000 this sequence version replaced gi:7523999.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0661A03
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-Primer ET; 92% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192742 bases at least Q40
Consensus quality: 194956 bases at least Q50
Consensus quality: 196287 bases at least Q20
Insert size: 203000; agarose-fp
Insert size: 199083; sum-of-contigs
Quality coverage: 4.81 in Q20 bases; agarose-fp
Quality coverage: 4.95 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1100: contig of 1100 bp in length
```

```
1101 1200: gap of unknown length
1201 2689: contig of 1489 bp in length
2690 2690: gap of unknown length
2690 3901: contig of 1112 bp in length
3902 4002: gap of unknown length
4002 5125: contig of 1124 bp in length
5125 5225: gap of unknown length
5225 6422: contig of 1197 bp in length
6422 6522: gap of unknown length
6522 7953: contig of 1431 bp in length
7953 8053: gap of unknown length
8053 9735: contig of 1682 bp in length
9735 9835: gap of unknown length
9835 11279: contig of 1444 bp in length
11279 11380: gap of unknown length
11380 12747: contig of 1368 bp in length
12747 12847: gap of unknown length
12847 15442: contig of 2595 bp in length
15442 15542: gap of unknown length
15542 17776: contig of 2234 bp in length
17776 20852: contig of 2976 bp in length
20852 20952: gap of unknown length
20952 22805: contig of 1853 bp in length
22805 22905: gap of unknown length
22905 24730: contig of 1825 bp in length
24730 24830: gap of unknown length
24830 26898: contig of 2068 bp in length
26898 26998: gap of unknown length
26998 29838: contig of 2840 bp in length
29838 29938: gap of unknown length
29938 33355: contig of 3417 bp in length
33355 33455: gap of unknown length
33455 36448: contig of 2993 bp in length
36448 36548: gap of unknown length
36548 38886: contig of 2338 bp in length
38886 38986: gap of unknown length
38986 43053: contig of 4067 bp in length
43053 43153: gap of unknown length
43153 47261: contig of 4108 bp in length
47261 47361: gap of unknown length
47361 50223: contig of 2862 bp in length
50223 50324: gap of unknown length
50324 54005: contig of 3682 bp in length
54005 54105: gap of unknown length
54105 58598: contig of 4494 bp in length
58598 58699: gap of unknown length
58699 65036: contig of 6337 bp in length
65036 65136: gap of unknown length
65136 68745: contig of 3609 bp in length
68745 74175: contig of 5330 bp in length
74175 74275: gap of unknown length
74275 78904: contig of 4629 bp in length
78904 79004: gap of unknown length
84470: contig of 5466 bp in length
84471 84571: gap of unknown length
84571 90144: contig of 5574 bp in length
90144 90244: gap of unknown length
90244 97102: contig of 6858 bp in length
97102 97202: gap of unknown length
97202 103353: contig of 6150 bp in length
103353 103453: gap of unknown length
103453 111622: contig of 8170 bp in length
111622 111722: gap of unknown length
111722 120658: contig of 8936 bp in length
120658 120758: gap of unknown length
120758 128603: contig of 7845 bp in length
128603 128704: gap of unknown length
128704 138599: contig of 9896 bp in length
138599 138600: gap of unknown length
138600 148106: contig of 9407 bp in length
148106 148206: gap of unknown length
```

FEATURES

148207 159001: contig of 10795 bp in length

* 159002 159101: gap of unknown length

* 159102 159183: contig of 10682 bp in length

* 169784 169883: gap of unknown length

* 169884 202983: contig of 33100 bp in length.

Location/Qualifiers

source

1. .202983

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="8"

/clone="RP11-661A3"

1. .1100

/note="assembly_name:Contig95"

misc_feature 1201. .2689

/note="assembly_name:Contig11"

misc_feature 2790. .3901

/note="assembly_name:Contig13"

misc_feature 4002. .5125

/note="assembly_name:Contig14"

misc_feature 5226. .6422

/note="assembly_name:Contig15"

misc_feature 6523. .7953

/note="assembly_name:Contig16"

misc_feature 8054. .9735

/note="assembly_name:Contig18"

misc_feature 9836. .11279

/note="assembly_name:Contig19"

misc_feature 11380. .12747

/note="assembly_name:Contig20"

misc_feature 12848. .15442

/note="assembly_name:Contig21"

misc_feature 15543. .17776

/note="assembly_name:Contig22"

misc_feature 17877. .20852

/note="assembly_name:Contig23"

misc_feature 20953. .22805

/note="assembly_name:Contig24"

misc_feature 22906. .24730

/note="assembly_name:Contig25"

misc_feature 24831. .26898

/note="assembly_name:Contig26"

misc_feature 26999. .29838

/note="assembly_name:Contig27"

misc_feature 29939. .33355

/note="assembly_name:Contig28"

misc_feature 33456. .36448

/note="assembly_name:Contig29"

misc_feature 36549. .38886

/note="assembly_name:Contig30"

misc_feature 38987. .43053

/note="assembly_name:Contig31"

misc_feature 43154. .47261

/note="assembly_name:Contig32"

misc_feature 47362. .50223

/note="assembly_name:Contig33"

misc_feature 50324. .54005

/note="assembly_name:Contig34"

misc_feature 54106. .58599

/note="assembly_name:Contig35"

misc_feature 58700. .65036

/note="assembly_name:Contig36"

misc_feature 65337. .68745

/note="assembly_name:Contig37"

misc_feature 68846. .74175

/note="assembly_name:Contig38"

misc_feature 74276. .78904

/note="assembly_name:Contig39"

misc_feature 79005. .84470

/note="assembly_name:Contig40"

misc_feature 84571. .90144

/note="assembly_name:Contig41"

misc_feature 90245. .97102

/note="assembly_name:Contig42"

Query Match

Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 ATTTTCATTTCTTGCTC 129

Db 93601 ATTTTCATTTCTTGCTC 93583

RESULT 34

AC019313 208618 bp DNA linear HTG 17-MAR-2000

LOCUS Homo sapiens chromosome 18 clone RP11-119P12 map 18, WORKING DRAFT

DEFINITION

SEQUENCE, 49 unordered pieces.

AC019313

AC019313.3 GI:7259732

VERSION HTG; HTGS PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 208618)

Homo sapiens chromosome 18, clone RP11-119P12

2 (bases 1 to 208618)

Unpublished

Birren, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barina, N., Beckert, R., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkert, G., Castle, A., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatae, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Melgrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vasiliyev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zody, M.

Direct Submission

Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 17, 2000 this sequence version replaced gi:6984442.

All repeats were identified using RepeatMasker:

Smith, A. F. A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MTBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L5386

Center clone name: 119 P 12

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 161318 bases at least Q40

Consensus quality: 181693 bases at least Q30

Consensus quality: 194050 bases at least Q20

Insert size: 203818; sum-of-contigs

Quality coverage: 3.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1052: contig of 1052 bp in length
* 1053 1152: gap of 100 bp
* 1153 2486: contig of 1334 bp in length
* 2487 2586: gap of 100 bp
* 2587 3719: contig of 1133 bp in length
* 3720 3819: gap of 100 bp
* 3820 3915: contig of 96 bp in length
* 3916 4015: gap of 100 bp
* 4016 5118: contig of 1103 bp in length
* 5119 5218: gap of 100 bp
* 5219 6275: contig of 1057 bp in length
* 6276 6375: gap of 100 bp
* 6376 7721: contig of 1346 bp in length
* 7722 7821: gap of 100 bp
* 7822 8956: contig of 1135 bp in length
* 8957 9056: gap of 100 bp
* 9057 10726: contig of 1670 bp in length
* 10727 10826: gap of 100 bp
* 10827 12216: contig of 1390 bp in length
* 12217 12316: gap of 100 bp
* 12317 13744: contig of 1428 bp in length
* 13745 13844: gap of 100 bp
* 13845 15089: contig of 1245 bp in length
* 15090 15189: gap of 100 bp
* 15190 16739: contig of 1550 bp in length
* 16740 16839: gap of 100 bp
* 16840 17920: contig of 1081 bp in length
* 17921 18020: gap of 100 bp
* 18021 19143: contig of 1123 bp in length
* 19144 19243: gap of 100 bp
* 19244 20489: contig of 1246 bp in length
* 20490 20589: gap of 100 bp
* 20590 21851: contig of 1262 bp in length
* 21852 21951: gap of 100 bp
* 21952 23358: contig of 1407 bp in length
* 23359 23458: gap of 100 bp
* 23460 25234: contig of 1776 bp in length
* 25235 25334: gap of 100 bp
* 25335 27169: contig of 1835 bp in length
* 27170 27269: gap of 100 bp
* 27270 29801: contig of 2532 bp in length
* 29802 32368: contig of 2367 bp in length
* 32369 32468: gap of 100 bp
* 32469 34415: contig of 2047 bp in length
* 34416 34515: gap of 100 bp
* 34516 36552: contig of 2037 bp in length
* 36553 36652: gap of 100 bp
* 36653 39620: contig of 2968 bp in length
* 39621 39720: gap of 100 bp
* 39721 42451: contig of 2731 bp in length
* 42452 42551: gap of 100 bp
* 42552 46217: contig of 3666 bp in length
* 46218 46317: gap of 100 bp
* 46318 48687: contig of 2370 bp in length
* 48688 48787: gap of 100 bp
* 48788 52204: contig of 3417 bp in length
* 52205 52304: gap of 100 bp
* 52305 55124: contig of 2820 bp in length
* 55125 55224: gap of 100 bp
* 55225 59236: contig of 4012 bp in length
* 59237 59336: gap of 100 bp
* 59337 63427: contig of 4091 bp in length
* 63428 63527: gap of 100 bp
* 63528 68235: contig of 4708 bp in length
* 68236 68335: gap of 100 bp

* 68336 74336: contig of 6001 bp in length
* 74337 74437 74437: gap of 100 bp
* 74438 79648: contig of 5212 bp in length
* 79649 79748: gap of 100 bp
* 79749 86527: contig of 6779 bp in length
* 86528 86627: gap of 100 bp
* 86628 93092: contig of 6465 bp in length
* 93093 93192: gap of 100 bp
* 93193 97736: contig of 4544 bp in length
* 97737 97836: gap of 100 bp
* 97837 103589: contig of 5753 bp in length
* 103590 103689: gap of 100 bp
* 103690 11389: contig of 7700 bp in length
* 11389 11489: gap of 100 bp
* 11490 117615: contig of 6126 bp in length
* 117616 117715: gap of 100 bp
* 117716 126146: contig of 8431 bp in length
* 126147 126246: gap of 100 bp
* 126247 133762: contig of 7516 bp in length
* 133763 133862: gap of 100 bp
* 133863 143183: contig of 9321 bp in length
* 143184 143283: gap of 100 bp
* 143284 154009: contig of 10726 bp in length
* 154010 154109: gap of 100 bp
* 154110 164224: contig of 10115 bp in length
* 164225 164324: gap of 100 bp
* 164325 176540: contig of 12216 bp in length
* 176541 176640: gap of 100 bp
* 176641 192181: contig of 15541 bp in length
* 192182 192281: gap of 100 bp
* 192282 208618: contig of 16337 bp in length.

FEATURES

source

1..208618
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-119P12"
/clone_1b="RP11-11 Human Male BAC"
1..1052
/note="assembly_fragment"
1153..2486
/note="assembly_fragment"
2587..3719
/note="assembly_fragment"
3820..3915
/note="assembly_fragment"
clone_end:T7
vector_side:right"
4016..5118
/note="assembly_fragment"
5219..6275
/note="assembly_fragment"
6376..7721
/note="assembly_fragment"
7822..8956
/note="assembly_fragment"
9057..10726
/note="assembly_fragment"
10827..12216
/note="assembly_fragment"
12317..13744
/note="assembly_fragment"
13845..15089
/note="assembly_fragment"
15190..16739
/note="assembly_fragment"
16840..17920

Query Match 8.3%; Score 19; DB 2; Length 208618;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Gaps 0;

DEFINITION Rattus norvegicus clone CH230-21A12, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC123450.3 GI:25086744
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 222692)
Muzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Angiano,D, Ayala-Bechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,D, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benamed,F, Biewald,K, Blair,D, Blankenburg,K, Blyth,P, Brown,M, Bryan,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davis,M,L, Davis,C, Day-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Haland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliy,K,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshuwa,L, Loulsegad,H, Lozado,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangun,B, Mapua,P, Martin,K, Martin,R, Martinez,E, McWhirre,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Mloasavilevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Norkovis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nsookumeh,O, Okwunnu,G, Olarnpungoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plopper,F, Poldexter,A, Popovic,D, Primus,E, Pu,L, L, Plazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajs,D, Sneed,A, Sodergren,E, Song,X-Z, Sorella,R, Sosa,J, Steime,W, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmali,K, Valas,R, Vera,Y, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczky,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausem,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R,A.
TITLE Rattus norvegicus clone CH230-21A12, WORKING DRAFT SEQUENCE, 2
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 222692)
Worley,K.C.
REFERENCE Direct Submission
TITLE Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL 3 (bases 1 to 222692)
REFERENCE Rat Genome Sequencing Consortium.
TITLE Direct Submission
AUTHORS Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

COMMENT BAYLOR PLAZA, HOUSTON, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:22856157.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXOG
Center clone name: CH230-21A12
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 212567 bases at least Q40
Consensus quality: 214333 bases at least Q30
Consensus quality: 215448 bases at least Q20
Estimated insert size: 217382; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 221265: contig of 221265 bp in length
* 221266 221365: gap of unknown length
* 221366 222692: contig of 1327 bp in length.

FEATURES
source
1..222692
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-21A12"
1..1611
/note="wgs contig"
misc_feature
1662..2265
/note="wgs contig"
misc_feature
/note="wgs contig"
BASE COUNT 57088 a 47682 c 48203 g 63526 t 6193 others
ORIGIN
Query Match 8.3%; Score 19; DB 2; Length 222692;
Best Local Similarity 100.0%; Pred.No.23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 TTTATAAGCAGTCATTT 27
|||||
Db 77617 TTTATAAGCAGTCATTT 77599

RESULT 37
AC127208 228638 bp DNA linear HTG 10-MAY-2003
AC127208/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-160KX0, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC127208
VERSION AC127208.3 GI:30522796
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

RESULT 38
AC023717/c
LOCUS
DEFINITION
AC023717
AC023717.3 GI:21930215
HTG: HTGS PHASE1; HTGS DRAFT; HTGS ACTIVEFIN.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 228802)
Muzny, D.M., Adams, C., Adio-Ochola, B., Ali-Osman, F.R., Allen, C.,
Alabrooks, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homes, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, U., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mathiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabat, K., Morgan, M., Morris, S.,
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkenko, S., Ogun, M., Okunodu, G.,
Oxague, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peterson, L., Pickens, R., Prime, E., Pul, L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Saverly, G.,
Scherer, S., Scott, G., Shen, H., Shoshitani, N., Slason, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Syarik, A., Taber, P., Tameris, A., Tameris, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Unpublished
2 (bases 1 to 228802)
Worley, K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 228802)
Worley, K.C.
Direct Submission
Submitted (23-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 23, 2002 this sequence version replaced gi:6997287.
Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: DRUG
Center clone name: RP98-46E23
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 254374 bases at least Q40
Consensus quality: 258931 bases at least Q30
Consensus quality: 271567 bases at least Q20
Estimated insert size: 222216; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 9 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 9673: contig of 9673 bp in length
* 9674 9773: gap of unknown length
* 9774 12080: contig of 2307 bp in length
* 12081 12180: gap of unknown length
* 12181 14308: contig of 2128 bp in length
* 14309 14408: gap of unknown length
* 14409 17229: contig of 2821 bp in length
* 17230 17329: gap of unknown length
* 17330 20215: contig of 2886 bp in length
* 20216 30315: gap of unknown length
* 30316 33623: contig of 13308 bp in length
* 33624 33723: gap of unknown length
* 33724 47696: contig of 13973 bp in length
* 47697 47796: gap of unknown length
* 47797 95023: contig of 47227 bp in length
* 95024 95123: gap of unknown length
* 95124 228802: contig of 133679 bp in length.
Location/Qualifiers
1..228802
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="X"
/clone="RP98-46E23"

BASE COUNT 64009 a 50361 c 49400 g 64230 t 802 others
ORIGIN
Query Match 8.3%; Score 19; DB 2; Length 228802;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Gy 175 ATTTTCAACCCGGGCTCAC 193
Db 163358 ATTTTCAACCCGGGCTCAC 163340

RESULT 39
AC096231
LOCUS
DEFINITION
AC096231.8 GI:30522338
HTG: HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 229077)

Wuzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhammed,F, Bismalo,K, Blair,J, Blankenburg,K, Blych,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Cente,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gbregregoris,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jollivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensuwa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheswari,M, Mahindartne,M, Mahmood,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, McWhiney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Mloasavjevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwankweli,O, Okwuonu,G, Olarunpogoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plopper,F, Polndexter,A, Popovic,D, Primus,B, Pu,L, L, Piazto,M, Quiroz,J, Rachin,E, Reeves,K, Regier,M,A, Reigh,R, Rially,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojase,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajic,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steinle,M, Strong,R, Sutton,A, Syatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villaseana,D, Walidron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczky,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Zoon,D, von Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R,A.

2 (bases 1 to 229077)

REFERENCE
AUTHORS

Unpublished

REFERENCE
AUTHORSREFERENCE
AUTHORS

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORSREFERENCE
AUTHORS

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GERM
Center clone name: CH230-33C23
----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 212493 bases at least Q40
Consensus quality: 215756 bases at least Q30
Consensus quality: 217252 bases at least Q20
Estimated insert size: 229406; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 225592: contig of 225592 bp in length
* 225593 225692: gap of unknown length
* 225693 229077: contig of 3385 bp in length.
Location/Qualifiers

FEATURES
source

1..229077
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-33C23"
misc_feature 1..1561
/note="wgs contig"

BASE COUNT 64800 a 42460 c 40825 g 70150 t 10842 others
ORIGIN

Query Match 8.3%; Score 19; DB 2; Length 229077;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TACTCAGAAATTTTGATG 47
Db 88115 TACTCAGAAATTTTGATG 88133

RESULT 40

AC095567/c

LOCUS

AC095567 229140 bp DNA linear HTG 09-MAY-2003
Rattus norvegicus clone CH230-7016, *** SEQUENCING IN PROGRESS ***,
7 unordered pieces.

ACCESSION

AC095567

VERSION

AC095567.6

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 229140)
Wuzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,

Baldwin D., Bandaruanaite D., Barber M., Bannestad M., Benahmed F.,
 Biswal J., Blair J., Blankenburg K., Blyth P., Brown M.,
 Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
 Cardenas J., Carter K., Cavazos I., Cesaar H., Center A.,
 Chacko V., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
 Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
 Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
 Delgado O., Denon S., Deramo C., Ding Y., Dinh H., Divya K.,
 Diaper H., Dugan-Rocha S., Dunn A., Durkin K., Duval B., Evans K.,
 Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
 Fernandez S., Finley M., Flagg N., Forbes L., Foster P.,
 Fraser C.M., Gabisi A., Gatta R., Garcia A., Garner T., Garza M.,
 Georgergis E., Geer K., Gill R., Grady M., Guerra W., Guerrero M.,
 Gunaratne P., Haaland W., Haml C., Hamilton C., Hamilton K.,
 Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
 Hernandez B., Hines S., Hladun S.L., Hodgson A., Hogues M.,
 Hollins B., Howells S., Huljak S., Hume J., Ideblid D., Jackson A.,
 Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
 Karpaty S., Kelly S., Khan Z., King L., Kovar C.,
 Kwis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
 Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
 Lorenehuwa L., Louiseged H., Lozado R.J., Lu X., Ma J.,
 Maheshwari M., Mahindartne M., Mahmoud M., Malloy M., Mangun A.,
 Mangus B., Mapua P., Martin K., Martin R., Martinez E.,
 Manthey S., McLeod M.P., McNelly T.Z., Meenen E.,
 Milosavljevic A., Miner G., Manja E., Montemayor J., Moore S.,
 Morgan M., Morris K., Morris S., Mundaasa M., Murphy M., Natr L.,
 Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
 Nwokilemeh O., Okwuonu G., Olarnpungsoon A., Pal S., Parks K.,
 Paternak S., Paul H., Perez A., Perez L., Pfannkuch C.,
 Plopper F., Poldexter A., Popovic D., Primus E., Pu L.-L.,
 Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
 Kelly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
 Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
 Sanders W., Savery G., Scherer S., Scott G., Shatman S., Shen H.,
 Shetty J., Shvartsbeyn A., Sison I., Sitter C.D., Snajs D.,
 Sneed A., Sodergren E., Song X.-Z., Sorrell R., Sosa J.,
 Steinle M., Strong R., Sutton A., Szatek A., Tabot P., Taylor C.,
 Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,
 Valas R., Varga V., Villasa D., Waldron L., Walker B., Wang J.,
 Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,
 Williams G., Wilson R., Wleciwyk R., Wooden H., Worley K.,
 Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
 Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von
 Niederhausen A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
 Weinstock G. and Gibbs R.A.
 Unpublished
 Direct Submission
 2 (bases 1 to 229140)
 Worley K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 229140)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 9, 2003 this sequence version replaced gi:24940786.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.tgsc.bcm.tmc.edu/Projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web Site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCSB
Center clone name: CH230-7016
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 211799 bases at least Q40
Consensus quality: 215054 bases at least Q30
Consensus quality: 217265 bases at least Q20
Estimated insert size: 219764; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) .
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 217989: contig of 217989 bp in length
* 217990 218089: gap of unknown length
* 218090 219464: contig of 1375 bp in length
* 219465 219564: gap of unknown length
* 219565 221481: contig of 1917 bp in length
* 221482 221581: gap of unknown length
* 221582 222864: contig of 1283 bp in length
* 222865 222964: gap of unknown length
* 222965 224313: contig of 1349 bp in length
* 224314 224413: gap of unknown length
* 224414 226184: contig of 1771 bp in length
* 226185 226284: gap of unknown length
* 226285 229140: contig of 2856 bp in length.
Location/Qualifiers
    1..229140
        /organism="Rattus norvegicus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10116"
        /clone="CH230-7016"
        /csize=42513..43616
misc_feature
    42513..43616
        /note="wgs_contig"
misc_feature
    75250..76439
        /note="wgs_contig"
misc_feature
    205968..207493
        /note="wgs_contig"
misc_feature
    210498..212309
        /note="wgs_contig"
BASE COUNT      65426 a 41824 c 42944 g   68932 t 10014 others
ORIGIN
Query Match          8.3%; Score 19; DB 2; Length 229140;
Beet Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative * 0; Mismatches 0; Indels 0; Gaps 0;
27 TCTACTCAGAAATTTTGA 45
|||||
|||||
|||||
AC113904              23163 bp       DNA         linear     HTG 19-NOV-2002
LOCUS               AC113904
DEFINITION          Rattus norvegicus clone CH230-318P21, *** SEQUENCING IN PROGRESS
AC113904
AC113904.5 GI:25072640
VERSION             HTG_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS

```

SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus

REFERENCE
AUTHORS 1 (bases 1 to 233163)
Nuzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biewald,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Bunay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabris,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gbrgeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M, Hollins,B, Howell,S, Huylk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowitz,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,U, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,D, Lorensuwa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, McWhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwaelelehu,O, Okwunnu,G, Olarunpasegun,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plopper,F, Polndexter,A, Popovic,D, Primus,E, Pu,L-L, Piazio,M, Quiroz,J, Rachin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojars,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,M, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajic,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,D, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczek,R, Woodson,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstock,G, and Gibbs,R,A.

TITLE Rattus norvegicus (Norway rat)
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 233163)
Worley,K,C.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233163)
TITLE Rat Genome Sequencing Consortium.
JOURNAL Direct Submission
AUTHORS Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 19, 2002 this sequence version replaced gi:23194968. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTRV
Center clone name: CH230-318P21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 209814 bases at least Q40
Consensus quality: 211069 bases at least Q30
Consensus quality: 211999 bases at least Q20
Estimated insert size: 220261; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 223189: contig of 223189 bp in length
* 223190 223289: gap of unknown length
* 223290 233163: contig of 9874 bp in length.
Location/Qualifiers
1. 233163
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-318P21"
1. 2826
/note="wgs_end_extension
clone_end:T7"
7180. 8127
/note="clone_boundary
clone_end:T7
site:
end_sequence: BZ141951"
BASE COUNT 66448 a 39633 c 40161 g 66269 t 20652 others
ORIGIN

Query Match 8.3%; Score 19; DB 2; Length 233163;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TTATTAAGCACTCAATTTC 28
|||||
Db 139093 TTATTAAGCACTCAATTTC 139111

RESULT 42
AC099082 235182 bp DNA linear HTG 10-MAY-2003
AC099082/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-11L9, WORKING DRAFT SEQUENCE, 7
unordere pieces.
AC099082
VERSION AC099082.5 GI:30522836
KEYWORDS HTG; HTGS_PHASEL; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 235182)
Allen, C., Allen, H., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Munty, D., Marie, D., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Ayala-Bechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, U., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davis, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Diexer, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, A., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgioyis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, M.,
Gundrat, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idlerbird, D., Jackson, A.,
Jackson, S., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kova, C.,
Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, H., Louisa, L., Louisa, H., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minig, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Narkewicz, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokileme, O., Okunolu, G., Olajunle, A., Palis, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poldre, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.,
Sanders, W., Savary, G., Scheer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smit, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhu, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL
TITLE
REFERENCE

2 (bases 1 to 235182)

Worley, K.C.

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235182)

Rat Genome Sequencing Consortium.

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23096197.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GCGO

Center clone name: CH230-1119

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 19951 bases at least Q40

Consensus quality: 205352 bases at least Q30

Consensus quality: 209335 bases at least Q20

Estimated insert size: 21211; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 223476: contig of 223476 bp in length

223477 223576: gap of unknown length

223577 224676: contig of 1100 bp in length

224677 224776: gap of unknown length

224777 225906: contig of 1130 bp in length

225907 226006: gap of unknown length

226007 227139: contig of 1133 bp in length

227140 227239: gap of unknown length

227240 228691: contig of 1452 bp in length

228692 228792: gap of unknown length

228793 229133: contig of 3122 bp in length

229134 232013: gap of unknown length

232014 235182: contig of 3169 bp in length.

Location/Qualifiers

1. 235182

/organism="Rattus norvegicus"

/db_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-1119"

/note="wgs contig"

misc_feature 1. 1778

BASE COUNT 67264 a 39463 c 41351 g 63457 t 23647 others

ORIGIN

Query Match 8.3%; Score 19; DB 2; Length 235182;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

29 TACTCAGAAATTTTGATG 47

Db 101737 TACTCAGAAATTTTGATG 101719

RESULT 43

AC126298 235700 bp DNA linear HTG 22-SEP-2002

LOCUS Rattus norvegicus clone CH230-31B17, *** SEQUENCING IN PROGRESS

DEFINITION *** 9 unordered pieces.

ACCESSION AC126298

BASE COUNT 55307 a 33066 c 32860 g 54709 t 59758 others
ORIGIN

Query Match 8.3%; Score 19; DB 2; Length 235700;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 TTATAAGCACTCAATTTC 28
|||||
Db 16206 TTATAAGCACTCAATTTC 16224

RESULT 44
AC108557
LOCUS
DEFINITION Rattus norvegicus clone CH230-106J14, *** SEQUENCING IN PROGRESS
AC108557 242803 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-106J14, ***
AC108557 3 unordered pieces.
AC108557.5 GI:30580652
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 242803)
Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryan, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falle, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, J., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Gante, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, R., Guevara, W.,
Gnarthe, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huily, S., Hume, J., Idebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, W., Louie, H., Lorado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahandarte, M., Mahmood, K., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minis, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwona, G., Olariupagoo, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pflanz, C.,
Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Rigge, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Silter, C. D., Smajic, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Sytek, A., Taber, Z., Usmani, K.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, P., Taylor, C.,
Vallas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G., and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 242803)
Worley, K. C.
Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 242803)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:33111081.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both ends sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GLEB
Center clone name: CH230-106J14
----- Summary Statistics
Assembly program: Atlas 3.0:
Consensus quality: 215425 bases at least Q40
Consensus quality: 217320 bases at least Q30
Consensus quality: 218537 bases at least Q20
Estimated insert size: 227859; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 238212: contig of 238212 bp in length
* 238213 238312: gap of unknown length
* 238313 241662: contig of 3350 bp in length
* 241663 241762: gap of unknown length
* 241763 242803: contig of 1041 bp in length.
location/Qualifiers
1. 242803
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-106J14"
1. 2264
/note="wgs end extension
clone end: T"
1550..2265
/note="clone boundary
clone end: T"
site: Ecoli

FEATURES
source
misc_feature
misc_feature
misc_feature

```

misc_feature      end_sequence:BH353942"
                  /note="clone boundary
                  106299..107098
                  clone_end:Sp6
                  site:EcORI
misc_feature      end_sequence:BH353944"
                  /note="wgs end extension
                  233686..235576
                  clone_end:Sp6"
misc_feature      235627..238212
                  /note="wgs end extension
                  clone_end:Sp6"
misc_feature      238313..239631
                  /note="wgs end extension
                  clone_end:Sp6"

BASE COUNT      67043 a 41721 c 41660 g 68976 t 23403 others
ORIGIN
Query Match      8.3%; Score 19; DB 2; Length 242803;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 TTATTAAGCACTCAATTTC 28
        |||||||
Db      21572 TTATTAAGCACTCAATTTC 21590

RESULT 45
AC097237/c      245082 bp      DNA      linear      HTG 10-MAY-2003
LOCUS      Rattus norvegicus clone CH230-63H19, WORKING DRAFT SEQUENCE.
DEFINITION      AC097237
ACCESSION      AC097237.6 GI:30522413
VERSION      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FUZZTOP.
KEYWORDS      Rattus norvegicus (Norway rat)
SOURCE      Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE      1 (bases 1 to 245082)
            Wuzny,D,Marite, Metzker,M, Lee, Abramzon,S, Adams,C, Alder,J,
            Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
            Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
            Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
            Bismalo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
            Bryant,N, Buhay,C, Burch,P, Cavazos,I, Ceasar,H, Centes,A,
            Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Centes,A,
            Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
            Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
            Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
            Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
            Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
            Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
            Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
            Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
            Gbregory,B, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W,
            Gunaratne,P, Haaland,W, Hamli,C, Hamilton,C, Hamilton,K,
            Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
            Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
            Hollins,B, Howells,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A,
            Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
            Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
            Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
            Liu,U, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
            Lorensbuhwa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J,
            Maheshwari,M, Mahindaratne,M, Mahmood,M, Malloy,K, Mangum,A,
            Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
            Mawhley,S, McLeod,M,P, McNeill,T,Z, Meenen,B,
            Mlosoajevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
            Morgan,M, Morris,K, Morris,S, Munidaa,M, Murphy,M, Nair,L,
            Nankervis,C, Neal,D, Newton,S, Nuyuen,N, Norris,S,
            Nwakoienh,O, Okwomu,G, Olarpunsaqon,A, Pal,S, Parks,K,
            Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkoch,C,

```

```

TITLE      Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,-L,
REFERENCE      Piazzi,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
AUTHORS      Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
TITLE      Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
JOURNAL      Sanders,W, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,
            Shetty,D, Shvartsbeyn,A, Sisson,L, Sitter,C,D, Smajs,D,
            Sneed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Soes,J,
            Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabour,P, Taylor,C,
            Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umani,K,
            Valae,R, Vera,V, Villana,D, Waldron,L, Walker,B, Wang,J,
            Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
            Williams,G, Willson,R, Wlezyk,R, Wooden,H, Worley,K,
            Wright,D, Wright,R, Wu,J, Yakub,S, Yen,P, Yoon,L, Yoon,V,
            Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhou,S, Dunn,D, von
            Niederhausern,A, Weles,R, Smith,D,R, Holt,R,A, Smith,H,O,
            Weinstein,G, and Gibbs,R,A.
            Direct Submission
            Unpublished
            2 (bases 1 to 245082)
            Worley,K,C.
            Direct Submission
            Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 245082)
            Rat Genome Sequencing Consortium.
            Direct Submission
            Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On May 10, 2003 this sequence version replaced gi:22855496.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GGSB
            Center clone name: CH230-63H19
            ----- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 23600 bases at least Q40
            Consensus quality: 23643 bases at least Q30
            Consensus quality: 238264 bases at least Q20
            Estimated insert size: 244959; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            1 245082: contig of 245082 bp in length.
            Location/Qualifiers
            1..245082

```

FEATURES
source

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-63H19"

misc_feature 1..1631 /note="wgs_contig"
BASE COUNT 64097 a 51810 c 53142 g 70982 t 5051 others
ORIGIN

Query Match 8.3%; Score 19; DB 2; Length 245082;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 TTATAAGCACTCATTT 27
Db 233568 TTATAAGCACTCATTT 233550

Search completed: February 4, 2004, 15:24:52
Job time : 3156 secs


```
PF 29-SEP-2000; 2000MO-CA01154.
XX
XX 29-SEP-1999; 99US-0156594.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
XX Buisse U, Chypre C, Fradet Y,
XX
XX WPI; 2001-258132/26.
XX
PT Novel nucleic acid encoding differentially expressed prostate cancer
PT antigen 3 mRNA containing additional sequence giving rise to long PCA3
PT mRNA, useful for diagnosis of mammal afflicted with prostate cancer -
XX
XX Claim 3(a); Fig 3; 60pp; English.
XX
CC The present sequence is that of an RT-PCR-amplified fragment of
CC human prostate cancer antigen 3 (PCA3) mRNA that includes an
CC additional 228 bp sequence, inserted between exons 3 and 4a. A 2nd
CC amplified fragment (see AAF30667) lacks this additional sequence.
CC The additional sequence interrupts the open reading frame of PCA3
CC protein, thereby yielding a truncated PCA3 protein. The shorter
CC form PCA3 RNA is associated with prostate cancer whereas the longer
CC form PCA3 RNA is associated with a non-malignant prostatic state,
CC such as benign prostatic hyperplasia. Based on the differential
CC expression of these 2 PCA3 RNA species, protocols for the diagnosis
CC of prostate disease are provided, including a method of diagnosing
CC the presence or predisposition to develop prostate cancer in a
CC patient. Also provided are therapeutic methods that use a nucleic
CC acid encoding a differentially expressed PCA3 mRNA molecule, an
CC antisense sequence, a protein encoded by a differentially expressed
CC PCA3 mRNA, or an antibody raised against such a protein.
XX
SQ Sequence 506 BP; 132 A; 123 C; 108 G; 143 T; 0 other;
Query Match 100.0%; Score 228; DB 22; Length 506;
Best Local Similarity 100.0%; Pred. No. 4.8e-108;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTAGTGGTTCTATTAAGACATCAATTTCTACTCAGAAATTTTATGGCCTTAGTTCT 60
Db 27 GTTAGTGGTTCTATTAAGACATCAATTTCTACTCAGAAATTTTATGGCCTTAGTTCT 86
QY 61 CTACTCGTTTCTATCTCTCTACTCACTGTCCTCCGGAATCACAACCGATTTTCTATT 120
Db 87 CTACTCGTTTCTATCTCTCTACTCACTGTCCTCCGGAATCACAACCGATTTTCTATT 146
QY 121 TCTTGCTCTGATATGTCGACTGCTCACTTGGAATTATCTCAGGAGTCTGGAATTTTC 180
Db 147 TCTTGCTCTGATATGTCGACTGCTCACTTGGAATTATCTCAGGAGTCTGGAATTTTC 206
QY 181 TACCGGGGCTCAGCTCGGTCCTCCATTTTGTCTCTCACTTTACAG 228
Db 207 TACCGGGGCTCAGCTCGGTCCTCCATTTTGTCTCTCACTTTACAG 254
RESULT 2
AAF30668/c
ID AAF30668 standard; cDNA; 20 BP.
XX
XX AAF30668;
AC
AC 11-JUN-2001 (first entry)
XX
XX Prostate cancer antigen 3 (PSA3) nucleic acid.
XX
XX PCA3; prostate cancer; antigen; marker; differential expression;
XX diagnosis; therapy; human; ss.
XX
XX Homo sapiens.
XX
XX WO200123550-A2.
XX
```

```
PD 05-APR-2001.
XX
XX 29-SEP-2000; 2000MO-CA01154.
PF
XX 29-SEP-1999; 99US-0156594.
PR
XX
XX (DIAG-) DIAGNOCURE INC.
XX
XX Buisse U, Chypre C, Fradet Y,
XX
XX WPI; 2001-258132/26.
XX
XX
XX Claim 6; Page 57; 60pp; English.
XX
XX The present sequence comprises nucleotides from human prostate
XX cancer antigen 3 (PCA3) long mRNA (see AAF30666). Claimed isolated
XX nucleic acid molecules consist of 10-50 nucleotides which
XX specifically hybridise to a differentially expressed long PCA3
XX mRNA, and are complementary to, or consist of, at least 10
XX consecutive nucleotides of the present sequence. Long PCA3 mRNA
XX includes a 228 bp sequence, inserted between exons 3a and 4, which
XX is absent in short PCA3 mRNA. Short PCA3 mRNA is associated with
XX prostatic cancer. Long PCA3 RNA is associated with a non-malignant
XX prostatic state. Differential expression of these 2 PCA3 RNA
XX species provides protocols for the diagnosis of prostate disease,
XX including a method of diagnosing the presence or predisposition to
XX develop prostate cancer in a patient. Also provided are
XX therapeutic methods that use a nucleic acid encoding a
XX differentially expressed PCA3 mRNA molecule, an antisense
XX sequence, a protein encoded by a differentially expressed PCA3
XX mRNA, or an antibody raised against such a protein.
XX
SQ Sequence 20 BP; 9 A; 1 C; 8 G; 2 T; 0 other;
Query Match 8.8%; Score 20; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 CGTTTCTATCTCTCTACTC 85
Db 20 CGTTTCTATCTCTCTACTC 1
RESULT 3
AAF30666/c
ID AAF30666 standard; cDNA; 506 BP.
XX
XX AAF30666;
AC
AC 11-JUN-2001 (first entry)
XX
XX Human differentially expressed PCA3 cDNA (long form).
XX
XX PCA3; prostate cancer; antigen; benign prostatic hyperplasia;
XX differential expression; diagnosis; gene therapy; chromosome 9;
XX human; ss.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX
XX FT exon 1..26
XX FT /*tag= a
XX FT /number= "3"
XX FT 27..254
XX FT /*tag= b
XX FT /note= "228 bp insertion"
XX FT 255..506
XX FT /*tag= c
XX FT /number= "4a"
XX
```

XX WO200123550-A2.
PN 05-APR-2001.
XX 29-SEP-2000; 2000WO-CA01154.
XX 29-SEP-1999; 99US-0156594.
XX (DIAG-) DIAGNOCURE INC.
XX Busse U, Chypre C, Fradet Y;
XX WPI; 2001-258132/26.
XX Novel nucleic acid encoding differentially expressed prostate cancer
PT antigen 3 mRNA containing additional sequence giving rise to long PCA3
PT mRNA, useful for diagnosis of mammal afflicted with prostate cancer -
XX
PS Claim 3(a); Fig 3; 60pp; English.
XX The present sequence is that of an RT-PCR-amplified fragment of
CC human prostate cancer antigen 3 (PCA3) mRNA that includes an
CC additional 228 bp sequence, inserted between exons 3 and 4a. A 2nd
CC amplified fragment (see AAF30667) lacks this additional sequence.
CC The additional sequence interrupts the open reading frame of PCA3
CC protein, thereby yielding a truncated PCA3 protein. The shorter
CC form PCA3 RNA is associated with prostate cancer whereas the longer
CC such as benign prostatic hyperplasia. Based on the differential
CC expression of these 2 PCA3 RNA species, protocols for the diagnosis
CC of prostate disease are provided, including a method of diagnosing
CC the presence or predisposition to develop prostate cancer in a
CC patient. Also provided are therapeutic methods that use a nucleic
CC acid encoding a differentially expressed PCA3 mRNA molecule, an
CC antisense sequence, a protein encoded by a differentially expressed
CC PCA3 mRNA, or an antibody raised against such a protein.
XX
SQ Sequence 506 BP; 132 A; 123 C; 108 G; 143 T; 0 other;
XX
Query Match 7.9%; Score 18; DB 22; Length 506;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 AGTGCCTTATTAAGCACT 21
DB 47 AGTGCCTTATTAAGCACT 30
XX
RESULT 4
AAS83274/c
ID AAS83274 standard; cDNA; 3414 BP.
XX
AC AAS83274;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #19078.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG19087.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 19078; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 3414 BP; 915 A; 806 C; 940 G; 753 T; 0 other;
XX
Query Match 7.9%; Score 18; DB 23; Length 3414;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 ATTAAGCACTCAATTCT 29
DB 1981 ATTAAGCACTCAATTCT 1964
XX
RESULT 5
ABV36501/c
ID ABV36501 standard; cDNA; 360 BP.
XX
AC ABV36501;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 36492.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
XX

PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 7541; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Table 1-9 (ABY00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient; the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (d) selecting a composition for inhibiting prostate cancer in a patient;
CC (e) assessing the prostate cell carcinogenic potential of a compound;
CC (f) determining whether prostate cancer has metastasized in a patient;
CC (g) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 360 BP; 104 A; 75 C; 77 G; 104 T; 0 other;
XX
Query Match 7.5%; Score 17; DB 23; Length 360;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 109 CGATTTCTATTCTTG 125
DB 270 CGATTTCTATTCTTG 254
XX
RESULT 6
ABN60692
ID ABN60692 standard; cDNA; 564 BP.
XX
AC ABN60692;
XX
XX 28-JUN-2002 (first entry)
XX
XX Human cancer related polynucleotide SEQ ID NO 659.
XX
XX Human, cytostatic; gene expression; gene mapping; tissue profiling;
XX gene therapy; cancer; tumour; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200214500-A2.
XX
XX 21-FEB-2002.
XX
XX 16-AUG-2001; 2001WO-US25840.
XX
XX 16-AUG-2000; 2000US-226326P.
XX
XX (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Escobedo J, Garcia PD, Suduth-Klinger J, Reinhard C, Randazzo F;
XX Lanson G, Scott EM, Zhang G, Kassem A, Pot D, Labat I;
XX WPI; 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially

PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth -
XX
XX Claim 1; SEQ ID NO 659; 883pp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
XX with cytostatic activity. The polynucleotide is used to produce a
XX polypeptide, to detect differentially expressed genes correlated with a
XX cancerous state of a mammalian cell and to inhibit tumour growth. The
XX polynucleotide is used as a probe in mapping and tissue profiling. The
XX encoded polypeptide and antibodies to the polypeptide can also be used
XX for therapeutic and diagnostic purposes. The polynucleotide is useful for
XX gene therapy.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 564 BP; 178 A; 76 C; 101 G; 209 T; 0 other;
XX
Query Match 7.5%; Score 17; DB 24; Length 564;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 TATTAAGCACTCAATTT 27
DB 202 TATTAAGCACTCAATTT 218
XX
RESULT 7
ABL22535
ID ABL22535 standard; DNA; 580 BP.
XX
AC ABL22535;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 19078.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-064150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 19078; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL20511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (AB57737-AB572072).
XX
XX The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
CC Sequence 580 BP; 141 A; 152 C; 144 G; 143 T; 0 other;
SQ
Query Match 7.5%; Score 17; DB 23; Length 580;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 CCGATTTCATTTCTT 124
DB 62 CCGATTTCATTTCTT 78
|||||
RESULT 8
ABX61876
ID ABX61876 standard; DNA; 601 BP.
XX
AC ABX61876;
XX
DT 25-FEB-2003 (first entry)
XX
DE Novel human transporter protein related polynucleotide #72.
XX
KM Human; gamma-aminobutyric acid; GABA; neurotransmitter transporter;
KM transgenic animal; ribozyme design; drug screening; gene therapy;
KM de.
XX
OS Homo sapiens.
XX
PN US2002142381-A1.
XX
PD 03-OCT-2002.
XX
PF 28-MAR-2001; 2001US-0818656.
XX
PR 28-MAR-2001; 2001US-0818656.
XX
PA (GONG/) GONG F.
PA (KERC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Gong F, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR WPI; 2003-102517/09.
XX
PT Novel human transporter protein, related to gamma-aminobutyric acid
PT neurotransmitter transporter subfamily useful as model for developing
PT human therapeutic targets and serves as target for human therapeutics -
XX
PS Disclosure; Page 76; 114pp; English.
XX
CC The invention describes an isolated human transporter peptide (I) that
CC is related to the gamma-aminobutyric acid (GABA) neurotransmitter
CC transporter subfamily. (I) and the nucleic acid encoding it (II) can be
CC used as models for the development of human therapeutic targets, aid in
CC the identification of therapeutic proteins and serve as targets for the
CC development of human therapeutic agents that modulate transporter
CC activity. (I) is used to raise antibodies or to elicit another immune
CC response, as a reagent in assays designed to quantitatively determine
CC levels of the protein in biological fluids, and as markers for tissues
CC in which the corresponding protein is preferentially expressed. The
CC transporter proteins isolated from humans and their human/mammalian
CC orthologues serve as targets for identifying agents for use in mammalian
CC therapeutic applications, and biological assays related to transporter
CC proteins that are related to members of the GABA neurotransmitter
CC transporter subfamily. The proteins and peptides also provide a target
CC for diagnosing a disease or predisposition to disease mediated by the
CC peptide, and are useful for treating a disorder characterized by absence
CC of, inappropriate, unwanted or altered expression of the protein. The
CC antibodies are also useful for assessing normal and aberrant subcellular
CC localisation of cells in various tissues in an organism, in

CC pharmacogenomic analysis, for tissue typing and for inhibiting protein
CC function. (II) is useful for constructing recombinant vectors, host cells
CC and transgenic animals; for designing ribozymes; in drug screening;
CC in diagnostic assays for qualitative changes in gene expression;
CC particularly in qualitative changes that lead to pathology; in gene
CC therapy; and to detect mutations in genes encoding transporters. This
CC sequence represents a polynucleotide related to the novel human
CC aminobutyric acid (GABA) transporter related protein of the invention.
XX
SQ Sequence 601 BP; 149 A; 145 C; 103 G; 203 T; 1 other;
QY
Query Match 7.5%; Score 17; DB 25; Length 601;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 94 TTCTATTTCTTGCTC 110
|||||
RESULT 9
ABX61877
ID ABX61877 standard; DNA; 601 BP.
XX
AC ABX61877;
XX
DT 25-FEB-2003 (first entry)
XX
DE Novel human transporter protein related polynucleotide #73.
XX
KM Human; gamma-aminobutyric acid; GABA; neurotransmitter transporter;
KM transgenic animal; ribozyme design; drug screening; gene therapy;
KM de.
XX
OS Homo sapiens.
XX
PN US2002142381-A1.
XX
PD 03-OCT-2002.
XX
PF 28-MAR-2001; 2001US-0818656.
XX
PR 28-MAR-2001; 2001US-0818656.
XX
PA (GONG/) GONG F.
PA (KERC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Gong F, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR WPI; 2003-102517/09.
XX
PT Novel human transporter protein, related to gamma-aminobutyric acid
PT neurotransmitter transporter subfamily useful as model for developing
PT human therapeutic targets and serves as target for human therapeutics -
XX
PS Disclosure; Page 76-77; 114pp; English.
XX
CC The invention describes an isolated human transporter peptide (I) that
CC is related to the gamma-aminobutyric acid (GABA) neurotransmitter
CC transporter subfamily. (I) and the nucleic acid encoding it (II) can be
CC used as models for the development of human therapeutic targets, aid in
CC the identification of therapeutic proteins and serve as targets for the
CC development of human therapeutic agents that modulate transporter
CC activity. (I) is used to raise antibodies or to elicit another immune
CC response, as a reagent in assays designed to quantitatively determine
CC levels of the protein in biological fluids, and as markers for tissues
CC in which the corresponding protein is preferentially expressed. The
CC transporter proteins isolated from humans and their human/mammalian
CC orthologues serve as targets for identifying agents for use in mammalian
CC therapeutic applications, and biological assays related to transporter
CC proteins that are related to members of the GABA neurotransmitter

CC transporter subfamily. The proteins and peptides also provide a target
CC for diagnosing a disease or predisposition to disease mediated by the
CC peptide, and are useful for treating a disorder characterized by absence
CC of, inappropriate, unwanted or altered expression of the protein. The
CC antibodies are also useful for assessing normal and aberrant subcellular
CC localization of cells in various tissues in an organism, in
CC pharmacogenomic analysis, for tissue typing and for inhibiting protein
CC function. (ii) is useful for constructing recombinant vectors, host cells
CC and transgenic animals for designing changes in gene expression,
CC in diagnostic assays for qualitative changes that lead to pathology, in gene
CC therapy, and to detect mutations in genes encoding transporters. This
CC sequence represents a polynucleotide related to the novel human
CC amino butyric acid (GABA) transporter related protein of the invention.
SQ Sequence 601 BP; 142 A; 151 C; 108 G; 199 T; 1 other;

Query Match 7.5%; Score 17; DB 25; Length 601;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 113 TTTCTATTCTTGCTC 129
Db 39 TTTCTATTCTTGCTC 55

RESULT 10
AA194586/C
ID AA194586 standard; cDNA; 773 BP.
XX
AC AA194586;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 661.
XX
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS Homo sapiens.
XX
PN WO20016719-A1.
XX
PD 13-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-JP01629.
XX
PR 07-MAR-2000; 2000JP-0159195.
XX
PA (CHIB-) CHIBA PREFECTURE.
XX
PA (HISM) HISAMITSU PHARM CO LTD.
XX
PI Nakagawara A;
XX
DR WPI; 2001-565584/63.
XX
PT Nucleic acids originating in gene expressed in human neuroblastoma,
PT useful as probe or primer in diagnosing prognosis of human
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT for anti-cancer agents
XX
PS Claim 1; Page 525; 2979pp; Japanese.
XX
CC The invention relates to novel genes (AA193926-AA197663) expressed in
CC human neuroblastoma. The nucleic acids are applicable as a probe or
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC susceptibility indicators or tumour markers for anti-cancer agents. The
CC gene information for diagnosing prognosis is related to factors similar
CC to that for N-myc and TrkA genes.
XX
SQ Sequence 773 BP; 250 A; 150 C; 120 G; 224 T; 29 other;

Query Match 7.5%; Score 17; DB 22; Length 773;
Best Local Similarity 100.0%; Pred. No. 68;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 21 TCAATTTCTACTGAA 37
Db 70 TCAATTTCTACTGAA 54

RESULT 11
AAH04021
ID AAH04021 standard; cDNA; 793 BP.
XX
AC AAH04021;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:856.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 856; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH1629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 793 BP; 232 A; 119 C; 157 G; 282 T; 3 other;

Query Match 7.5%; Score 17; DB 22; Length 793;

Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 CTCACCTGGATTATCC 161
DB 220 CTCACCTGGATTATCC 236

RESULT 12

AAS53792/C
ID AAS53792 standard; DNA; 927 BP.

AC AAS53792;

DT 13-FEB-2002 (first entry)

DE Helicobacter pylori DNA for cellular proliferation protein #246.

KM Antisense; ds; prokaryotic cellular proliferation gene;

KM antibiotic; antibacterial; drug design.

OS Helicobacter pylori.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haseelbeck R, Ohlsen KL, Zysek JD, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR P-PSDB; AAU35933.

PS Claim 27; Seq ID No 7429; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Streptococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 927 BP; 314 A; 176 C; 196 G; 241 T; 0 other;

XX Query Match 7.5%; Score 17; DB 23; Length 927;

Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTTGC 126
DB 19 GATTTCTATTCTTGC 3

RESULT 13

AAV37398
ID AAV37398 standard; DNA; 1252 BP.

AC AAV37398;

DT 13-OCT-1998 (first entry)

DE Streptococcus pneumoniae coding region.

XX coding region; ORF; open reading frame; antibacterial;

XX infection; prevention; meningitis; ss.

OS Streptococcus pneumoniae.

PN Key Location/Qualifiers

FT CDS 1030..1251

FT /tag= a

FT /product= putative gamma-glutamyl phosphate reductase

XX WO9819689-A1.

PD 14-MAY-1998.

XX 27-OCT-1997; 97NO-US19226.

XX 01-NOV-1996; 96US-0029930.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;

PI Reid RH, Zarfos PN;

DR WPI; 1998-286586/25.

DR P-PSDB; AAW61010.

PS Claim 1; Page 104-105; 130pp; English.

XX The sequence is that of a coding region isolated from

CC S. pneumoniae. Its encoded protein, or agonists of it,

CC may be useful as an antibacterial for treatment or

CC prevention of infection, specifically caused by S. pneumoniae

CC (particularly meningitis) but possibly also Helicobacter

CC pylori (ulcers and gastric cancer). It may be of particular

CC use before insertion of an in-dwelling device or any other

CC invasive procedure. The protein, or nucleic acid encoding

CC it, can also be used in vaccines to induce a cellular

CC and/or humoral immune response, or to screen for other

CC antibacterials. The DNA may also contain flanking sequences

CC that are potential sources of control elements for bacterial

CC gene expression. Detecting a sequence encoding the protein

CC can be used diagnostically, e.g. to detect a mutation for

CC serotyping or classifying infectious agents.

XX Sequence 1252 BP; 351 A; 235 C; 331 G; 335 T; 0 other;

XX Query Match 7.5%; Score 17; DB 19; Length 1252;

XX Best Local Similarity 100.0%; Pred. No. 68;

XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 ATTCTACTCAGAAATT 40
|||||
DB 1153 ATTCTACTCAGAAATT 1169

RESULT 14
ABX06622 standard; DNA; 1260 BP.
XX
AC ABX06622;
XX
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #910.
XX
KM Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KM ear infection; antiinflammatory; antibacterial; immunostimulant;
KM auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae type 4 strain.
XX
PN WO200277021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-1B02163.
XX
PR 27-MAR-2001; 2001GB-0007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Maeignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
DR P-PSDB; ABU01335.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection -
XX
XX
PS Claim 6; SEQ ID No 1819; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the pairs of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2489 identified coding region from the genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1260 BP; 348 A; 238 C; 340 G; 334 T; 0 other;
XX

Query Match 7.5%; Score 17; DB 25; Length 1260;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 ATTCTACTCAGAAATT 40
|||||
DB 1165 ATTCTACTCAGAAATT 1181

RESULT 15
AAD31177
ID AAD31177 standard; DNA; 1357 BP.
XX
AC AAD31177;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human delta-5-desaturase gene control region.
XX
XX
KM Human; delta-5-desaturase; D5D; drug screening; lipid metabolism;
KM diabetic neuropathy; arterial hypertension; hypercholesterolaemia;
KM atherosclerotic heart disease; chronic inflammatory disorder;
KM autoimmune disorder; allergic eczema; atopic disorder;
KM rheumatoid arthritis; lymphocyte proliferation; natural killer cell;
KM T-cell; cytotoxicity; macrophage; chemotaxis; antigen presentation;
KM major histocompatibility class II; pro-inflammatory cytokine;
KM interleukin; tumour necrosis factor; adhesion molecule; eczema;
KM psoriasis; acute respiratory distress syndrome; ARDS;
KM articular cartilage degradation; ACD; cancer; ds.
XX
XX
OS Homo sapiens.
XX
PN WO200234940-A2.
XX
PD 02-MAY-2002.
XX
PF 26-OCT-2001; 2001WO-CA01520.
XX
PR 26-OCT-2000; 2000US-243009P.
XX
PA (XENO-) XENON GENETICS INC.
XX
PI Winther MD, Knickle LC, Haardt M, Allen SJ, Ponton A;
PI De Antueno RJ, Jenkins DK, Nwaka SO;
XX
DR WPI; 2002-454647/48.
XX
PT Novel control region of delta-5-desaturase gene useful as a target for
PT screening compounds useful in the treatment of diseases involving
PT abnormal lipid metabolism including diabetic neuropathy -
XX
XX

Claim 1; Fig 1; 93pp; English.

The invention relates to human delta-5-desaturase (hD5D) gene control
region used as a target for screening drugs useful in treatment of
diseases involving abnormal lipid metabolism including diabetic
neuropathy, arterial hypertension, hypercholesterolaemia, atherosclerotic
heart disease, chronic inflammatory disorders, autoimmune disorders,
allergic eczema and other atopic disorders, inflammatory processes such
as rheumatoid arthritis, diminished lymphocyte proliferation,
T-cell-mediated cytotoxicity, natural killer cell activity,
macrophage-mediated cytotoxicity, monocyte and neutrophil chemotaxis,
major histocompatibility class II expression and antigen presentation,
production of pro-inflammatory cytokines (interleukins 1 and 6, tumour
necrosis factor) and adhesion molecule expression, eczema, psoriasis,
acute respiratory distress syndrome (ARDS), articular cartilage
degradation (ACD) and cancer. Host cell containing the control
region of D5D gene is useful for screening for a modulator capable
of regulating the expression of a mammalian D5D gene, especially for


```
CC identifying modulators that modulate lipid metabolism or diabetic
CC neuropathy. The screening method is an assay for identifying modulators
CC that modulate the n-3 lipid metabolic pathway, conversion of
CC 18:3n3-22:6n3, or n-9 lipid metabolic pathway, conversion of
CC 16:0-22:4n9 or n-6 lipid metabolic pathway, conversion of 18:2n6-22:5n6.
CC The human D5d control region provides a powerful tool for dissecting the
CC role of D5d gene expression and inducing modifications, which eliminate
CC or control alterations associated with metabolic disorders. The
CC present sequence is human delta-5-desaturase (hD5d) gene control region.
XX
SQ Sequence 1357 BP; 335 A; 417 C; 344 G; 261 T; 0 other;

Query Match      7.5%; Score 17; DB 24; Length 1357;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 GTAGAGCTTATAAG 17
Db      266 GTAGAGCTTATAAG 282

RESULT 16
AAC51976
ID AAC51976 standard; DNA; 1675 BP.
XX
XX AAC51976;
AC
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 70130.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
```

```

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147203.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

```

```

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 7.5%; Score 17; DB 21; Length 1675;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 9 TTTATTAAGCCTCCACT 25
Db 846 TTTATTAAGCCTCCACT 862

```

```

RESULT 17
ABV77738/C
ID ABV77738 standard; cDNA; 1842 BP.
XX
AC ABV77738;
XX
DT 03-FEB-2003 (first entry)
XX
DE Human zinc finger protein 10.56 coding sequence.
XX
KW Human; zinc finger protein 10.56; tumour; nervous system disease;
KW haemopathy; development disturbance; HIV infection; cytostatic; anti-HIV;
KW gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CD5 1269..1559
FT /*tag= a
FT /product= "zinc finger protein 10.56"
XX
XX CN1345803-A.
XX
PD 24-Apr-2002.
XX
PF 26-SEP-2000; 2000CN-0125422.
XX
PR 26-SEP-2000; 2000CN-0125422.
XX
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2002-539319/58.
XX
XX P-PSDB; ABP59031.
XX
XX Novel polypeptide-human zinc protein 10.56 and polynucleotide encoding
XX the polypeptide -
XX
XX Claim 6; Page 25-26 (Disclosure); 33pp; Chinese.
XX
XX The present sequence is the coding sequence for human zinc finger protein
XX 10.56. The protein is useful for treating several diseases, such as solid
XX tumours, nervous system disease, haemopathy, development disturbance and
XX HIV infection.
XX
XX Sequence 1842 BP; 516 A; 394 C; 369 G; 563 T; 0 other;
SQ

```

```

Query Match 7.5%; Score 17; DB 24; Length 1842;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 205 CATATTGCTCCTCACT 221
Db 104 CATATTGCTCCTCACT 88

```

```

RESULT 18
ABN69051

```

ID ABN69051 standard; DNA; 2070 BP.
 XX
 AC ABN69051;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Streptococcus polynucleotide SEQ ID NO 6015.
 XX
 KM Streptococcus GAS; GAS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KM anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-00282727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraeier C;
 PI Tettein H;
 XX
 DR WPI; 2002-352536/38.
 DR P-PSDB; ABP28420.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 7; Page 3763; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 2070 BP; 751 A; 373 C; 381 G; 565 T; 0 other;
 Query Match 7.5%; Score 17; DB 24; Length 2070;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 CTTTATTAAGCACTCAA 24
 Db 1363 CTTTATTAAGCACTCAA 1379
 RESULT 19
 AAX08433 standard; DNA; 2264 BP.
 XX
 AC AAX08433;

XX
 DT 28-JUN-1999 (first entry)
 XX
 DE Catalase gene of human kidney cells.
 XX
 KM Manganese containing superoxide dismutase; MnSOD; IDDM;
 KM diabetes mellitus; treatment; therapy; nitric oxide; NO; beta cell;
 KM fatty acid; lipotoxic; cytotoxic; cytokine; osteoporosis;
 KM inflammatory disease; autoimmune disease; neurodegenerative disease;
 KM ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 71..1654
 FT /tag= a
 FT /product= "Catalase"
 XX
 PN WO9906059-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 30-JUL-1998; 98WO-US15781.
 XX
 PR 03-MAR-1998; 98US-0055092.
 PR 30-JUL-1997; 97US-0055092.
 XX
 PA (BETA-) BETAGENE INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Clark SA, Hohmeier H, Koyama K, Lee Y, Newgard CB;
 PI Ohneda M, Shimabukuro, Thigpen A, Unger RH;
 XX
 DR WPI; 1998-153448/13.
 DR P-PSDB; AAM96321.
 XX
 PT Protection of mammalian cells against immunotoxicity or lipotoxicity
 PT - used for treating, e.g. diabetes, obesity, wasting syndromes,
 PT osteoporosis, inflammatory diseases, autoimmune diseases or
 PT neurodegenerative diseases
 XX
 PS Disclosure; Page 241-242; 253pp; English.
 XX
 CC Inhibition of cytokine mediated immunotoxicity of cells can be
 CC achieved by blocking free radical production or the accumulation of
 CC free radicals in that cell. Treatment of insulin dependent diabetes
 CC mellitus (IDDM) can be achieved by by blocking nitric oxide (NO)
 CC production in a pancreatic beta cell and by providing a composition
 CC comprising an agent that reduces levels of fatty acids in the cells
 CC and protects beta-cells of the subject against lipid-mediated cell
 CC death. Cells can also be protected against nitric oxide mediated
 CC cytotoxicity by introducing into the cell an antioxidant agent.
 CC The methods can be used for protecting cells against immunotoxicity
 CC mediated by, e.g. IL-1 beta, IL-1 alpha, gamma IFN, TNF alpha, TNF
 CC beta, IL-8, IL-2, IL-6, IL-2, IL-3, IL-5, IL-7, IL-9, IL-14, IL-17,
 CC granulocyte-macrophage colony stimulating factor or monocyte
 CC chemottractant protein-1. The methods can be used for the treatment
 CC of e.g. insulin-dependent diabetes mellitus (IDDM), NIDDM, obesity,
 CC wasting syndromes, short stature, osteoporosis, inflammatory
 CC diseases, autoimmune diseases, or neurodegenerative diseases.
 XX
 SQ Sequence 2264 BP; 634 A; 511 C; 509 G; 610 T; 0 other;
 Query Match 7.5%; Score 17; DB 20; Length 2264;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 TCTACTCAGAAATTTT 43
 Db 2047 TCTACTCAGAAATTTT 2063
 RESULT 20

AAA75733
ID AAA75733 standard; cDNA; 2279 BP.
XX
AC AAA75733;
XX
DT 22-JAN-2001 (first entry)
XX
DE Nucleotide sequence of catalase cDNA.
XX
KW Adeno-associated virus; catalase; superoxide dismutase; demyelination;
KW optic nerve; reactive oxygen species; optic neuritis; optic disk edema;
KW demyelinating disease; allergic encephalomyelitis; multiple sclerosis;
KW allergic encephalomyelitis; blood brain barrier; ss.
XX
OS Synthetic.
XX
PN WO200054595-A1.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-US06839.
XX
PR 15-MAR-1999; 99US-0124398.
XX
PA (GUY/) GUY J.
PA (GIX/) OI X.
PA (HAUS/) HAUSWIRTH W W.
XX
PI Guy J, OI X, Hauswirth W W;
XX
DR WPI: 2000-618854/59.
XX
PT Use of recombinant adeno-associated virus composition for treating
PT demyelinating disorders e.g. multiple sclerosis and allergic
PT encephalomyelitis, comprises viral constructs expressing catalase or
PT superoxide dismutase -
XX
PS Disclosure; Page 74-75; 78pp; English.
XX
CC The specification describes the use of a recombinant adeno-associated
CC virus (rAAV) composition. The rAAV comprises a polynucleotide sequence
CC encoding a mammalian catalase or superoxide dismutase polypeptide to
CC reduce demyelination in an optic nerve. The rAAV are useful for
CC scavenging reactive oxygen species in a mammal suspected of having
CC optic neuritis, to reduce the effects of demyelinating disease in a
CC mammal and prophylactically used to suppress blood brain barrier
CC disruption in a mammal having identified allergic encephalomyelitis.
CC The rAAV composition is useful for reducing symptoms associated with
CC demyelinating diseases such as optic neuritis, multiple sclerosis,
CC allergic encephalomyelitis, where the symptoms associated with these
CC disease include optic disk edema, increase of optic nerve cell count,
CC disruption of blood brain barrier integrity, increased levels of
CC hydrogen peroxide and demyelination of axons. The present sequence
CC is expressed using the rAAV of the invention.
XX
SQ Sequence 2279 BP; 642 A; 514 C; 512 G; 611 T; 0 other;
XX
Query Match 7.5%; Score 17; DB 21; Length 2279;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 TCTACTCGAGAAATTTT 43
DB 2044 TCTACTCGAGAAATTTT 2060

XX
DE Pain regulated cDNA sequence 75.
XX
KW Pain; analgesic; gene therapy; neurological disorder;
KW neurodegenerative disease; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200212338-A2.
XX
PD 14-FEB-2002.
XX
PF 03-AUG-2001; 2001WO-EP09011.
XX
PR 03-AUG-2000; 2000DE-1037759.
XX
PA (CHEF) GRUENENTHAL GMBH.
XX
PI Gillen C, Wetzel I, Wnendt S, Weine E, Schaefer MK;
XX
DR WPI: 2002-257469/30.
XX
PR P-SDB; ABB85028.
XX
PT Identifying pain-regulating compounds, useful for treating chronic pain
PT and for diagnosis, by measuring binding of compounds to specific
PT peptides and proteins -
XX
PS Claim 1; Fig 41; 213pp; German.
XX
CC The invention relates to identifying pain-regulating substances (A)
CC comprising (i) incubating a test substance with a cell (or preparation
CC from it) that has synthesised a peptide or protein (B) and (ii) measuring
CC either binding of the test substance to (B) or some functional parameter
CC that is altered by this binding. The method is useful for identifying
CC pain-regulating substances (A) with analgesic activity (A) along with
CC nucleic acid (ABU8411-ABU8441) that encode proteins (B),
CC ABB8506-ABB85037) that interact with (A); (B); vectors containing the
CC nucleic acid; antibodies against (B); cells that express (B) and agents
CC that bind to (B); are all useful for treating pain, particularly chronic
CC pain, including use in gene therapy. The same materials can also be used
CC for diagnosis, e.g. of neurological and neurodegenerative diseases. The
CC present sequence is that of a polynucleotide of the invention.
XX
SQ Sequence 2279 BP; 642 A; 514 C; 512 G; 611 T; 0 other;
XX
Query Match 7.5%; Score 17; DB 24; Length 2279;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 TCTACTCGAGAAATTTT 43
DB 2044 TCTACTCGAGAAATTTT 2060

RESULT 22
AA160810 standard; cDNA; 2309 BP.
XX
AC AA160810;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4799.
XX
KW Human; nociceptive; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX

PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM41654.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 4799; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2309 BP; 654 A; 521 C; 517 G; 617 T; 0 other;
 XX
 Query Match 7.5%; Score 17; DB 22; Length 2309;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 TCTACTCAGAAATTTT 43
 Db 2071 TCTACTCAGAAATTTT 2087
 RESULT 23
 AA159024
 ID AA159024 standard; cDNA; 2318 BP.
 XX
 AC AA159024;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1227.
 XX
 KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX Homo sapiens.
 OS
 XX
 PN WO200153312-A1.
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM39868.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1227; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2318 BP; 655 A; 529 C; 522 G; 612 T; 0 other;
 XX
 Query Match 7.5%; Score 17; DB 22; Length 2318;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 TCTACTCAGAAATTTT 43
 Db 2078 TCTACTCAGAAATTTT 2094
 RESULT 24
 AAK52557
 ID AAK52557 standard; cDNA; 2652 BP.
 XX
 AC AAK52557;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 2086.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation; ss.
 XX Homo sapiens.
 OS
 XX WO200157190-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 05-FEB-2001; 2001WO-US04098.
 PF
 XX 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR P-PSDB; AAM79424.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 4467; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 SQ Sequence 2652 BP; 707 A; 635 C; 635 G; 675 T; 0 other;
 XX
 QY Query Match 7.5%; Score 17; DB 22; Length 2652;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 27 TCTACTCAGAAATTTT 43
 2414 TCTACTCAGAAATTTT 2430
 XX
 RESULT 25
 AAK51573
 ID AAK51573 standard; cDNA; 2661 BP.
 XX
 AC AAK51573;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 118.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.

XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR P-PSDB; AAM78440.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 765-767; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 SQ Sequence 2661 BP; 713 A; 638 C; 637 G; 673 T; 0 other;
 XX
 QY Query Match 7.5%; Score 17; DB 22; Length 2661;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 27 TCTACTCAGAAATTTT 43
 2415 TCTACTCAGAAATTTT 2431
 XX
 RESULT 26
 ABL22534
 ID ABL22534 standard; DNA; 2662 BP.
 XX
 AC ABL22534;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19075.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.

PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW,
XX WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 19075; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2662 BP; 748 A; 596 C; 604 G; 714 T; 0 other;
XX
Query Match 7.5%; Score 17; DB 23; Length 2662;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 CCGATTTCCTATTTCTT 124
DB 1062 CCGATTTCCTATTTCTT 1078
XX
RESULT 27
AB122059
ID ABL22059 standard; DNA; 5790 BP.
XX
AC ABL22059;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 17650.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW,
XX WPI; 2001-656860/75.
XX
DR

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 17650; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5790 BP; 1480 A; 1401 C; 1414 G; 1495 T; 0 other;
XX
Query Match 7.5%; Score 17; DB 23; Length 5790;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 171 CTGGATTTCCTACCCG 187
DB 3598 CTGGATTTCCTACCCG 3614
XX
RESULT 28
AAS30407
ID AAS30407 standard; DNA; 6352 BP.
XX
AC AAS30407;
XX
DT 21-NOV-2001 (first entry)
XX
DE DNA encoding novel prostate gland antigen, Seq ID No 265.
XX
KM Human; noctropic; neuroprotective; cytostatic; antiapoptosis;
KM antianaemic; dermatological; immunosuppressive; antiinflammatory;
KM antiallergic; antirheumatic; virucide; hepatotropic; nephrotropic;
KM osteopathic; prostate gland; prostaticitis; adenocarcinoma; hair loss;
KM prostaticitis; malacoplakia; adenocarcinoma; benign prostatic hypertrophy;
KM hyperplasia; carcinoma; prostate neoplastic disorder; skin aging;
KM reproductive system disorder; autoimmune disorder; urinary system;
KM systemic lupus erythematosus; rheumatoid arthritis; cardiovascular;
KM blood-related disorder; hyperproliferative disorder; respiratory;
KM neurological disorder; endocrine disorder; inflammatory disorder;
KM liver disorder; wound healing; food preservative; ds.
XX
XX Homo sapiens.
OS
XX WO20015447-A1.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01330.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214986.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218299.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220966.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227189.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229511.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0231243.
PR 06-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249289.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-476223/51.
Novel isolated prostate gland related polypeptide useful for diagnosis
and treatment of disorders of prostate such as prostatic dysplasia,
prostatitis, prostaticitis, benign prostatic hypertrophy and malacoplakia
-
Claim 1; SEQ ID No 265; 512pp; English.
The invention relates to novel isolated prostate gland related nucleic
acids (I) and polypeptides (II). (I) and (II) are useful for diagnosis,
prognosis, prevention, and/or treatment of diseases and/or disorders of
the prostate such as acute non-bacterial prostatitis, chronic non-
bacterial prostatitis, acute bacterial prostatitis, prostatic dysplasia,
prostatitis, granulomatous prostatitis, malacoplakia, benign prostatic
hypertrophy or hyperplasia, and prostate neoplastic disorders, including
adenocarcinomas, transitional cell carcinomas, ductal carcinomas, and
squamous cell carcinomas. (I), (II) and antibody to (II) are useful for

CC diagnosing and treating reproductive system disorders (Paget's disease),
CC autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC arthritis), blood-related disorders (sickle cell anaemia),
CC hyperproliferative disorders, urinary system disorders
CC (glomerulonephritis), cardiovascular disorders (arrhythmias),
CC respiratory disorders, musculoskeletal system disorders, neutral activity
CC and neurological disorders (Alzheimer's disease and Parkinson's disease),
CC endocrine disorders (Addison's disease), gastrointestinal disorders
CC (inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC pancreatic and gall bladder disorders, disorders of the large intestine,
CC developmental and inherited disorders, diseases at the cellular level,
CC and wound healing and epithelial cell proliferation. (I) or (II) is
CC useful to prevent skin aging, for preventing hair loss, to maintain

Query Match 7.5%; Score 17; DB 22; Length 6352;

Best Local Similarity 100.0%; Pred. No. 67;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ATTTTGTATGGCCTTAA 54
|||||

Db 924 ATTTTGTATGGCCTTAA 940

RESULT 29
AAL04434
ID AAL04434 standard; DNA; 6352 BP.

XX AAL04434;

XX 21-NOV-2001 (first entry)

DE Human reproductive system related antigen DNA SEQ ID NO: 7122.

XX Human; reproductive system related antigen; reproductive system disorder;

KM cancer; gene therapy; ds.

XX Homo sapiens.

PN WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236358.
PR 29-SEP-2000; 2000US-0236359.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246522.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246603.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.

Query Match 7.5%; Score 17; DB 22; Length 18657;

Best Local Similarity 100.0%; Pred No. 67; Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 TTTCTATTCTTGCCCT 128
Db 6324 TTTCTATTCTTGCCCT 6340

RESULT 33

ID AAK84863 standard; DNA; 23527 BP.

AAK84863;

AC AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

XX AAK84863;

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0234984.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249300.
 PR 17-NOV-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-483426/52.
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS Disclosure; SEQ ID NO 39675; 3071pp + Sequence listing; English.
 PS
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytotoxic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 XX Sequence 23527 BP; 6788 A; 5335 C; 5423 G; 5981 T; 0 other;

Query Match 7.5%; Score 17; DB 22; Length 23527;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 ATGGCCTTAAGTTCCTC 61
|||
Db 4475 ATGGCCTTAAGTTCCTC 4491

RESULT 34
AAK84865
ID AAK84865 standard; DNA; 23527 BP.
XX AAK84865;
AC
XX
XX 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39677.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
PN MO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226866.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241821.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249247.
PR 17-NOV-2000; 2000US-0249248.
PR 17-NOV-2000; 2000US-0249249.
PR 17-NOV-2000; 2000US-0249250.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250161.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251990.
PR 06-DEC-2000; 2000US-0251991.
PR 08-DEC-2000; 2000US-0251992.
PR 08-DEC-2000; 2000US-0251993.
PR 08-DEC-2000; 2000US-0251994.
PR 08-DEC-2000; 2000US-0251995.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX Disclosure; SEQ ID NO 39677; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK4702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 23527 BP; 6788 A; 5333 C; 5424 G; 5982 T; 0 other;
Query Match 7.5%; Score 17; DB 22; Length 23527;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 ATGGCCTTAAGTTCCTC 61
Db 4475 ATGGCCTTAAGTTCCTC 4491
RESULT 35
AAK84864
ID AAK84864 standard; DNA; 23533 BP.
XX
AC AAK84864;
XX
DT 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39676.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM

KM cytoetic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
XX
XX W0200157182-A2.
PN
XX
PD 09-AUG-2001.
XX
PF 17-JUN-2001; 2001MO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227809.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
XX

PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251719.
PR 06-DEC-2000; 2000US-0251729.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251988.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 39676; 3071pp + Sequence listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87594 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 23533 BP; 6796 A; 5332 C; 5425 G; 5980 T; 0 other;

Query Match 7.5%; Score 17; DB 22; Length 23533;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 ATGGCCTTAAGTTCCTC 61
Db 4474 ATGGCCTTAAGTTCCTC 4490

RESULT 36
AAV52155
ID AAV52155 standard; DNA; 28171 BP.
AC AAV52155;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
XX
PN MO9818931-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97MO-US19588.
XX 31-OCT-1996; 96US-0029960.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
DR
XX Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
PS Claim 1; Page 273-289; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridize to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 28171 BP; 8507 A; 5080 C; 6440 G; 8144 T; 0 other;
Query March 7.5%; Score 17; DB 19; Length 28171;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 ATTTCTACTCAGAAATT 40
Db 6998 ATTCTACTCAGAAATT 7014
RESULT 37
AAL36829/c
ID AAL36829 standard; DNA; 28313 BP.
XX
AC AAL36829;
XX
DT 08-JAN-2002 (first entry)
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3194.
XX
KW Cytoskeletal; immunosuppressive; noctropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; de.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241806.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251988.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-451937/48.
 DR
 XX
 PT Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 PS Example 2, SEQ ID NO 3194; 781bp + Sequence Listing; English.
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (AB03087-AB0304109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human

CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pot_sequences.
 XX
 SQ Sequence 28313 BP; 6388 A; 6515 C; 7079 G; 8331 T; 0 other;
 Query Match 7.5%; Score 17; DB 22; Length 28313;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 75 CCTTCTACTACTGTC 91
 Db 6228 CCTTCTACTACTGTC 6212
 RESULT 38
 ABX59817/c
 ID ABX59817 standard; cDNA; 28313 BP.
 XX
 AC ABX59817;
 XX
 DT 26-FEB-2003 (first entry)
 DE cDNA encoding novel human musculoskeletal system antigen #2161.
 XX
 KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;
 KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;
 KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
 KW post-operative tissue repair; limb regeneration; neuronal growth;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW AIDS-related complex; chondrocyte growth; bone regeneration;
 KW periodontal regeneration; tissue transport; bone graft; skin aging;
 KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
 KW cell growth; organ transplant; cell differentiation; body height;
 KW weight; hair colour; eye colour; skin; percentage of adipose tissue;
 KW pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;
 KW depression; tendency for violence; pain; reproductive capability;
 KW hormone level; endocrine level; appetite; libido; memory; stress;
 KW storage capability; fat content; lipid content; protein content;
 KW carbohydrate content; vitamin content; cofactor content;
 KW nutritional component.
 XX
 OS Homo sapiens.
 XX
 PN US2002147140-A1.
 XX
 PD 10-OCT-2002.
 XX
 PF 17-JAN-2001; 2001US-0764877.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.

PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 12-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234979P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239335P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
XX (ROSE/) ROSEN C A.
XX (RUBE/) RUBEN S M.
XX (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-128199/12.
XX
XX Isolated nucleic acid molecules encoding musculoskeletal system
XX associated polypeptides, useful for detecting disorders, e.g. cancer -
XX
XX Disclosure; SEQ ID NO 3194; 321pp; English.
XX
XX The invention describes an isolated nucleic acid molecule comprising a
XX sequence encoding musculoskeletal system associated polypeptides useful
XX for detecting disorders, e.g., cancer or cancer metastases, in animals
XX or humans. The nucleic acid; stimulates re-vascularisation of ischaemic
XX tissues associated with conditions such as thrombosis, arteriosclerosis,
XX and other cardiovascular conditions; treats wounds due to injuries,
XX burns, post-operative tissue repair, and ulcers; stimulates angiogenesis
XX and limb regeneration; stimulates neuronal growth; can treat and prevent
XX neuronal damage occurring in certain disorders or neurodegenerative
XX conditions, such as, Alzheimer's disease, Parkinson's disease, and
XX AIDS-related complex; stimulates chondrocyte growth, thus they can be
XX used to enhance bone and periodontal regeneration and aid in tissue
XX transports or bone grafts; prevents skin aging due to sunburn by
XX stimulating keratinocyte growth; prevents hair loss, since FGF family
XX members activate hair-forming cells and promotes melanocyte growth;
XX stimulates growth and differentiation of hematopoietic cells and bone
XX marrow cells when used in combination with other cytokines; maintains
XX organs before transplantation or for supporting cell culture of primary
XX tissues; induces tissue of mesodermal origin to differentiate in early
XX embryos; increases or decreases the differentiation or proliferation of

CC embryonic stem cells, besides, hematopoietic lineage; modulates
CC mammalian characteristics, such as, body height, weight, hair colour, eye
CC colour, skin, percentage of adipose tissue, pigmentation, size, and shape
CC (e.g., cosmetic surgery); modulates mammalian metabolism; changes
CC mammal's metal state or physical state by influencing biorythms
CC circadian rhythms, depression, tendency for violence, tolerance for pain,
CC reproductive capabilities, hormonal or endocrine levels, appetite,
CC libido, memory, or stress; increases or decreases storage capabilities,
CC fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors
CC or other nutritional components. This sequence encodes a novel human
CC musculoskeletal system antigen.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID=20020147140.
XX
SQ Sequence 28313 BP; 6388 A; 6515 C; 7079 G; 8331 T; 0 other;
Query Match 7.5%; Score 17; DB 25; Length 28313;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 75 CCTTCCCTACTGCTGC 91
Db 6228 CCTTCCCTACTGCTGC 6212
RESULT 39
ID AAL36833 standard; DNA; 29228 BP.
XX
XX AAL36833;
AC
XX 08-JAN-2002 (first entry)
DT
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3198.
DE
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
XX
XX Homo sapiens.
XX
XX
XX W0200155367-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01338.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 14-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226661.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233065.
 PR 14-SEP-2000; 2000US-0233066.
 PR 21-SEP-2000; 2000US-0234222.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-451937/48.
 XX
 XX
 PT isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 XX
 PS Example 2; SEQ ID NO 3198; 781pp + Sequence listing; English.
 XX
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (AAB03087-AAB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the

CC from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID=20020147140.
XX
SQ Sequence 29228 BP; 6450 A; 6876 C; 7498 G; 8404 T; 0 other;
Query Match 7.5%; Score 17; DB 25; Length 29228;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 CCTTCCTACTGCTGTC 91
DB 3774 CCTTCCTACTGCTGTC 3758

RESULT 41
AB18664/c
ID AB18664 standard; DNA; 30143 BP.
XX
AC AB18664;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7465.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li FWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1: SEQ ID NO 7465; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB186176-AB186511), expressed DNA
CC sequences (AB1861840-AB186175) and the encoded proteins
CC (AB1861737-AB1861702).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 30143 BP; 8018 A; 6821 C; 6422 G; 8882 T; 0 other;

Query Match 7.5%; Score 17; DB 23; Length 30143;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 CTACCGATTTCGATTT 121
DB 11695 CTACCGATTTCGATTT 11679

RESULT 42
ABX61804
ID ABX61804 standard; DNA; 40645 BP.
XX
AC ABX61804;
XX
DT 25-FEB-2003 (first entry)
XX
DE Genomic DNA encoding novel human GABA transporter related protein.
XX
KW Human; gamma-aminobutyric acid; GABA; neurotransmitter transporter;
KM transgenic animal; ribozyme design; drug screening; gene therapy;
KM gene; ds.
XX
OS Homo sapiens.
XX
PN US2002142381-A1.
XX
PD 03-OCT-2002.
XX
PF 28-MAR-2001; 2001US-0818656.
XX
PR 28-MAR-2001; 2001US-0818656.
XX
PA (GONG/) GONG F.
PA (KETCHUM) KETCHUM K A.
PA (DPRR/) DI FRANCESCO V.
PA (BEASLEY) BEASLEY E M.
XX
PI Gong F, Ketchum KA, Di Francesco V, Beasley EW;
XX
DR WPI; 2003-102517/09.
XX
PT Novel human transporter protein, related to gamma-aminobutyric acid
PT neurotransmitter transporter subfamily useful as model for developing
PT human therapeutic targets and serves as target for human therapeutics -
XX
PS Claim 4; Fig 3; 114pp; English.
XX
CC The invention describes an isolated human transporter peptide (I) that
CC is related to the gamma-aminobutyric acid (GABA) neurotransmitter
CC transporter subfamily. (I) and the nucleic acid encoding it (II) can be
CC used as models for the development of human therapeutic targets, aid in
CC the identification of therapeutic proteins and serve as targets for the
CC development of human therapeutic agents that modulate transporter
CC activity. (I) is used to raise antibodies or to elicit another immune
CC response, as a reagent in assays designed to quantitatively determine
CC levels of the protein in biological fluids, and as markers for tissues
CC in which the corresponding protein is preferentially expressed. The
CC transporter proteins isolated from humans and their human/mammalian
CC orthologues serve as targets for identifying agents for use in mammalian
CC therapeutic applications, and biological assays related to transporter
CC proteins that are related to members of the GABA neurotransmitter
CC transporter subfamily. The proteins and peptides also provide a target
CC for diagnosing a disease or predisposition to disease mediated by the
CC peptide, and are useful for treating a disorder characterized by absence
CC of, inappropriate, unwanted or altered expression of the protein. The
CC antibodies are also useful for assessing normal and aberrant subcellular
CC localisation of cells in various tissues in an organism. In
CC pharmacogenomic analysis, for tissue typing and for inhibiting protein
CC function. (II) is useful for constructing recombinant vectors, host cells
CC and transgenic animals; for designing ribozymes; in drug screening;
CC in diagnostic assays for qualitative changes in gene expression,
CC particularly in qualitative changes that lead to pathology; in gene
CC therapy; and to detect mutations in genes encoding transporters. This
CC sequence encodes a novel human aminobutyric acid (GABA) transporter
CC related protein.
XX
SQ Sequence 40645 BP; 9831 A; 9356 C; 9734 G; 11724 T; 0 other;

Query Match 7.5%; Score 17; DB 25; Length 40645;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 TTCTATTCTGCTC 129
 Db 36001 TTCTATTCTGCTC 36017

RESULT 43
 ID ABR83575/c
 XX ABR83575 standard; cDNA; 136284 BP.
 AC ABR83575;
 XX
 XX 14-AUG-2002 (first entry)
 DE Human cDNA differentially expressed in granulocytic cells #146.
 XX
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.
 XX
 XX MO200228999-A2.
 PN
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US0821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 XX (GENE-) GENE LOGIC INC.
 PA
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 XX WPI; 2002-435328/46.
 DR
 XX
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 PS Claim 1; SEQ ID NO 146; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA; preferably in an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,

CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 136284 BP; 41693 A; 28668 C; 28000 G; 37923 T; 0 other;
 SQ

Query Match 7.5%; Score 17; DB 24; Length 136284;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 TCTACTCAGAAATTTT 43
 Db 3635 TCTACTCAGAAATTTT 3619

RESULT 44
 ID ABR10147
 XX ABR10147 standard; cDNA; 185035 BP.
 AC ABR10147;
 XX
 XX 04-DEC-2002 (first entry)
 DE Human breast cancer associated coding sequence SEQ ID NO: 281.
 XX
 KW Human; breast specific gene; breast cancer; differential expression;
 KW cytostatic; gene therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX MO200259271-A2.
 PN
 XX
 PD 01-AUG-2002.
 XX
 PF 25-JAN-2002; 2002WO-US02176.
 XX
 PR 25-JAN-2001; 2001US-263757P.
 PR 25-APR-2001; 2001US-286090P.
 PR 23-MAY-2001; 2001US-292517P.
 XX
 XX (GENE-) GENE LOGIC INC.
 PA
 PI Orr MS, Naton M, Digians JC, Zeng W;
 XX
 XX WPI; 2002-674803/72.
 DR
 XX
 XX
 PT Diagnosing breast cancer in a patient comprises detecting the level of
 PT gene expression in cell or tissue samples, where a differential gene
 PT expression is indicative of breast cancer -
 XX
 PS Claim 1; SEQ ID NO 281; 260pp + Sequence Listing; English.

XX The present invention relates to methods of diagnosing breast cancer in a
 CC patient, which comprise detecting the level of expression in a tissue
 CC sample of two or more genes selected from those shown in ABR09867-
 CC ABR1112, where a differential expression of the genes indicates breast
 CC cancer. The methods are useful in diagnosing, treating, detecting the
 CC progression, and in monitoring treatment of breast cancer in patients.
 CC The methods are also useful as a screening tool for agents that modulate
 CC the onset or progression of breast cancer. The breast cancer genes may be
 CC used as diagnostic markers for the prediction or identification of the
 CC malignant state of breast tissue, for confirming the type and progression
 CC of cancer, and for drug screening and assays. The present sequence is a
 CC coding sequence of the invention.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIP0
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 185035 BP; 42256 A; 51727 C; 51210 G; 39842 T; 0 other;

Query Match 7.5%; Score 17; DB 24; Length 185035;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTAGTCTTATAAG 17
 |||||
 Db 62467 GTAGTCTTATAAG 62483

RESULT 45

AA543104
 ID AA543104 standard; DNA; 325791 BP.

AA543104;

18-DEC-2001 (first entry)

Human Oestrogen receptor beta gene.

XX Human; Oestrogen receptor beta; ERbeta; ds; SNP; chromosome 6q.25.1;
 XX single nucleotide polymorphism; cardiovascular disease;
 KM autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;
 KM osteoarthritis; osteoporosis; breast cancer; endometrial cancer.

OS Homo sapiens.

MO200162793-A2.

30-AUG-2001.

20-FEB-2001; 2001WO-US05360.

22-FEB-2000; 2000US-0183755.

24-JAN-2001; 2001US-0768185.

(PEKE) PE CORP NY.

Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;

WPI; 2001-582041/65.

P-PSDB; AAU27322.

Oestrogen receptor gene and protein polymorphisms useful for diagnosis

of individuals at risk of developing bone disorders -

Example 2; Figure 1; 245bp; English.
 CC The invention relates to a novel isolated peptide comprising or
 CC consisting of an amino acid sequence selected from an amino acid sequence
 CC of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
 CC 10 amino acids), antibodies against them, nucleic acids encoding
 CC them (including vectors for transforming cells). The gene for human
 CC ERbeta is located on chromosome 6q.25.1. The variants are encoded
 CC by single nucleotide polymorphisms (SNP). The variant peptides and
 CC proteins can be used in assays to determine the biological
 CC activity of the protein, to raise antibodies, as a reagent in assays
 CC designed to quantitatively determine levels of the protein in
 CC biological fluids, to identify compounds that modulate receptor
 CC activity and to screen compounds for the ability to stimulate or
 CC inhibit interaction between the receptor protein and a target molecule
 CC that normally interacts with the receptor protein e.g. oestrogen.
 CC The antibody can be used to isolate the protein, to assess expression in
 CC disease states e.g. cardiovascular disease and autoimmune disease (e.g.
 CC systemic lupus erythematosus, arthritis, rheumatism and osteoarthritis),
 CC osteoporosis, breast cancer and endometrial cancer. In addition
 CC the antibodies can be used in pharmacogenomic analysis and inhibiting
 CC protein function, e.g. blocking the binding of the oestrogen receptor
 CC protein to a binding partner such as a ligand. The nucleic acids

CC encoding the proteins can be used as probes, primers, chemical
 CC intermediates and in biological assays. The present sequence is the
 CC human ERbeta gene.

XX Sequence 325791 BP; 94098 A; 68292 C; 67970 G; 95431 T; 0 other;

Query Match 7.5%; Score 17; DB 22; Length 325791;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

45 ATGGCCTTAGTCCCTC 61
 |||||
 Db 165554 ATGGCCTTAGTCCCTC 165570

Search completed: February 4, 2004, 15:29:44
 Job time : 291 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 14:31:59 ; Search time 70 Seconds
(Without alignments)
1437.648 Million cell updates/sec

Title: US-09-675-650-1_COPY_27_254

Perfect score: 228
Sequence: 1 gtaagtcctataaacac.....ttgtccctccattcacag 228

Scoring table: OLIGO_NTC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUG.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	7.5	1896	US-09-107-532A-248	Sequence 248, App
2	17	7.5	2264	US-09-126-109-9	Sequence 9, Appl
3	17	7.5	2318	US-09-620-312D-916	Sequence 916, App
4	17	7.5	28171	US-08-961-527-22	Sequence 22, Appl
5	16	7.0	300	US-09-313-284A-6133	Sequence 6133, Ap
6	16	7.0	439	US-09-702-705-1614	Sequence 1614, Ap
7	16	7.0	439	US-09-736-457-1614	Sequence 1614, Ap
8	16	7.0	493	US-09-702-705-13	Sequence 13, Appl
9	16	7.0	493	US-09-736-457-13	Sequence 13, Appl
10	16	7.0	516	US-09-702-705-46	Sequence 46, Appl
11	16	7.0	516	US-09-736-457-46	Sequence 46, Appl
12	16	7.0	970	US-09-702-705-794	Sequence 794, App
13	16	7.0	970	US-09-736-457-794	Sequence 794, App
14	16	7.0	1152	US-09-205-258-67	Sequence 67, Appl
15	16	7.0	1819	US-09-453-702B-188	Sequence 188, App
16	16	7.0	6154	US-08-504-617-1	Sequence 1, Appl
17	16	7.0	7015	US-09-177-249-6	Sequence 6, Appl
18	16	7.0	38653	US-09-922-445-1	Sequence 1, Appl
19	16	7.0	319608	US-09-539-333D-1	Sequence 1, Appl
20	16	7.0	319608	US-09-679-409-1	Sequence 1, Appl
21	15	6.6	920	US-09-221-017B-254	Sequence 254, App
22	15	6.6	957	US-09-252-991A-51	Sequence 51, Appl
23	15	6.6	975	US-09-107-532A-1108	Sequence 1108, Ap
24	15	6.6	1059	US-09-784-508-16	Sequence 16, Appl
25	15	6.6	1077	US-09-784-508-3	Sequence 3, Appl
26	15	6.6	1491	US-09-328-352-2461	Sequence 2461, Ap
27	15	6.6	1611	US-09-697-367-7	Sequence 7, Appl

28	15	6.6	1886	US-08-936-165A-224	Sequence 224, App
29	15	6.6	2042	US-08-933-821-16	Sequence 16, Appl
30	15	6.6	2042	US-08-934-494-5	Sequence 5, Appl
31	15	6.6	2042	US-08-960-507-16	Sequence 16, Appl
32	15	6.6	2042	US-09-143-068-5	Sequence 5, Appl
33	15	6.6	2042	US-09-143-707-5	Sequence 5, Appl
34	15	6.6	2042	US-09-202-089-5	Sequence 5, Appl
35	15	6.6	2042	US-09-136-828-16	Sequence 16, Appl
36	15	6.6	2042	US-09-332-928A-16	Sequence 16, Appl
37	15	6.6	2042	US-09-511-133-5	Sequence 5, Appl
38	15	6.6	2042	US-09-136-801-16	Sequence 16, Appl
39	15	6.6	2042	US-09-332-929-16	Sequence 16, Appl
40	15	6.6	2042	US-09-690-169-5	Sequence 5, Appl
41	15	6.6	2042	US-09-511-631-5	Sequence 5, Appl
42	15	6.6	2042	US-09-333-075-16	Sequence 16, Appl
43	15	6.6	2042	US-09-690-189-5	Sequence 5, Appl
44	15	6.6	2042	US-09-202-088A-16	Sequence 16, Appl
45	15	6.6	2042	US-09-333-077-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-107-532A-248
; Sequence 248, Application US/09107532A
; Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 248:
SEQUENCE CHARACTERISTICS:
LENGTH: 1896 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...1896

SEQUENCE DESCRIPTION: SEQ ID NO: 248:
US-09-107-532A-248

Query Match 7.5%; Score 17; DB 4; Length 1896;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 TTTCTATTCTGCT 128
|||||
539 TTTCTATTCTGCT 555

RESULT 2

US-09-126-109-9
Sequence 9, Application US/09126109
Patent No. 6171856
GENERAL INFORMATION:
APPLICANT: Thigpen, Anice
APPLICANT: Hohmeier, Hans-Ewald
APPLICANT: Newgard, Christopher B.
APPLICANT: Ungert, Roger H.
APPLICANT: Shimabukuro, Michio
APPLICANT: Chen, Guaxun
APPLICANT: Rhodes, Christopher J.
APPLICANT: Hugl, Sigrun R.
APPLICANT: Cousin, Sharon
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TO NO-MEDIATED CYTOTOXICITY
TITLE OF INVENTION: TO NO-MEDIATED CYTOTOXICITY
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESS: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/126,109
FILING DATE: 30-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,092
FILING DATE: 30-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US Unknown
FILING DATE: 03-MAR-1998
ATTORNEY/AGENT INFORMATION:
NAME: McMillian, Nadeela R.
REGISTRATION NUMBER: P-43,363
REFERENCE/DOCKET NUMBER: UTS-560
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2264 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-126-109-9

Query Match 7.5%; Score 17; DB 3; Length 2264;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 27 TCTACTCAGAAATTTT 43
|||||

DB 2047 TCTACTCAGAAATTTT 2063

RESULT 3

US-09-620-312D-916
Sequence 916, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dunru
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghaast
APPLICANT: Dmanac, Radoje T.
TITLE OF INVENTION: Polypeptides
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP28
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: PC_FL_genes Version 1.0
SEQ ID NO 916
LENGTH: 2318
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (103)..(1686)
US-09-620-312D-916

Query Match 7.5%; Score 17; DB 4; Length 2318;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TCTACTCAGAAATTTT 43
|||||
DB 2078 TCTACTCAGAAATTTT 2094

RESULT 4

US-08-961-527-22
Sequence 22, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2

```
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 28171 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-22
```

```
Query Match          7.5%; Score 17; DB 4; Length 28171;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      24 ATTCTACTCAGAAATT 40
      |||||
Db      6998 ATTCTACTCAGAAATT 7014
```

```
RESULT 5
US-09-313-294A-6323/C
Sequence 6323, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Laijudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 6323
LENGTH: 300
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6476212 700351604H1
LOCATION: 92, 254, 257, 294
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-6323
```

```
Query Match          7.0%; Score 16; DB 4; Length 300;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      122 CTGCGCTGTATTGTC 137
      |||||
Db      144 CTGCGCTGTATTGTC 129
```

```
RESULT 6
US-09-702-705-1614/C
Sequence 1614, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
```

```
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1614
LENGTH: 439
TYPE: DNA
ORGANISM: Homo sapiens
US-09-702-705-1614
```

```
Query Match          7.0%; Score 16; DB 4; Length 439;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      212 GTCCTCCACTTTCACA 227
      |||||
Db      193 GTCCTCCACTTTCACA 178
```

```
RESULT 7
US-09-736-457-1614/C
Sequence 1614, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1614
LENGTH: 439
TYPE: DNA
ORGANISM: Homo sapiens
US-09-736-457-1614
```

```
Query Match          7.0%; Score 16; DB 4; Length 439;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      212 GTCCTCCACTTTCACA 227
      |||||
Db      193 GTCCTCCACTTTCACA 178
```

```
RESULT 8
US-09-702-705-13
Sequence 13, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
```

```
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedavick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FaastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 493
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(493)
OTHER INFORMATION: n = A,T,C or G
US-09-702-705-13
```

```
Query Match
Best Local Similarity 100.0%; DB 4; Length 493;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 212 GTCCTCCACTTTCACA 227
Db 413 GTCCTCCACTTTCACA 428
```

```
RESULT 9
US-09-736-457-13
Sequence 13, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedavick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FaastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 493
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(493)
OTHER INFORMATION: n = A,T,C or G
US-09-736-457-13
```

```
Query Match
Best Local Similarity 100.0%; DB 4; Length 493;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 212 GTCCTCCACTTTCACA 227
Db 413 GTCCTCCACTTTCACA 428
```

```
RESULT 10
US-09-702-705-46
Sequence 46, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedavick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FaastSeq for Windows Version 3.0
SEQ ID NO 46
LENGTH: 516
TYPE: DNA
ORGANISM: Homo sapien
US-09-702-705-46
```

```
Query Match
Best Local Similarity 100.0%; DB 4; Length 516;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 212 GTCCTCCACTTTCACA 227
Db 413 GTCCTCCACTTTCACA 428
```

```
RESULT 11
US-09-736-457-46
Sequence 46, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedavick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FaastSeq for Windows Version 3.0
SEQ ID NO 46
LENGTH: 516
TYPE: DNA
ORGANISM: Homo sapien
US-09-736-457-46
```

```
Query Match
Best Local Similarity 100.0%; DB 4; Length 516;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 212 GTCCTCCACTTTCACA 227
Db 413 GTCCTCCACTTTCACA 428
```

RESULT 12
US-09-702-705-794/C
; Sequence 794, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 794
; LENGTH: 970
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-794

Query Match 7.0%; Score 16; DB 4; Length 970;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 GTCTCCACTTCA 227
Db 223 GTCTCCACTTCA 208

RESULT 13
US-09-736-457-794/C
; Sequence 794, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aljun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 794
; LENGTH: 970
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-736-457-794

Query Match 7.0%; Score 16; DB 4; Length 970;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 GTCTCCACTTCA 227
Db 223 GTCTCCACTTCA 208

RESULT 14
US-09-205-258-67/C
; Sequence 67, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 67
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (668)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (745)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1015)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1088)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1110)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1113)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-205-258-67

Query Match          7.0%; Score 16; DB 4; Length 1152;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      190 TCACCTCGTCCCTCC 205
Db      501 TCACCTCGTCCCTCC 486

RESULT 15
US-09-453-702B-186/c
; Sequence 188, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: NO. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-Dec-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296,95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1819
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 188:
US-09-453-702B-188
```

```

Query Match          7.0%; Score 16; DB 4; Length 1819;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      111 ATTTCTATTCTTCTGC 126
Db      350 ATTTCTATTCTTCTGC 335

RESULT 16
US-08-504-617-1/c
; Sequence 1, Application US/08504617
; Patent No. 6521236
; GENERAL INFORMATION:
; APPLICANT: Willems, Martha Jacoba
; TITLE OF INVENTION: Vector vaccines of recombinant
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 6521236el
; STREET: 1330 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/504,617
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/211,150
; FILING DATE: March 22, 1994
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6154 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Feline herpesvirus (FHV-1)
STRAIN: G2620
IMMEDIATE SOURCE:
CLONE: Lambda FHV04
FEATURE:
NAME/KEY: CDS
LOCATION: 127..1281
OTHER INFORMATION: /label= ORF-1
FEATURE:
NAME/KEY: CDS
LOCATION: 1460..3058
OTHER INFORMATION: /label= ORF-2
FEATURE:
NAME/KEY: CDS
LOCATION: 3055..3357
OTHER INFORMATION: /label= ORF-3
FEATURE:
NAME/KEY: CDS
LOCATION: 3505..3963
OTHER INFORMATION: /label= ORF-4
FEATURE:
NAME/KEY: CDS
LOCATION: complement (4256..4897)
OTHER INFORMATION: /label= ORF-5
FEATURE:
NAME/KEY: CDS
LOCATION: complement (5138..6142)
OTHER INFORMATION: /label= ORF-6
US-08-504-617-1
Query Match
Best Local Similarity 7.0%; Score 16; DB 4; Length 6154;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 105 CTACCGATTTCTATT 120
Db 3256 CTACCGATTTCTATT 3241
RESULT 17
US-09-177-249-6
Sequence 6, Application US/09177249
GENERAL INFORMATION:
APPLICANT: Fischer, Robert L.
APPLICANT: Ohad, Nir
APPLICANT: Kiyosue, Tomohiro
APPLICANT: Yadegari, Ramon
APPLICANT: Margossian, Linda
APPLICANT: Harada, John
APPLICANT: Goldberg, Robert B.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
FILE REFERENCE: 023070-086120US
CURRENT APPLICATION NUMBER: US/09/177,249
CURRENT FILING DATE: 1998-10-22
EARLIER APPLICATION NUMBER: US 09/071,838
EARLIER FILING DATE: 1998-05-01
NUMBER OF SEQ ID NOS: 324
```

```
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 7015
TYPE: DNA
ORGANISM: Arabidopsis sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(7014)
OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1)
FEATURE:
NAME/KEY: CDS
LOCATION: (2)..(7015)
OTHER INFORMATION: fertilization-independent endosperm 1 (FIE1)
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(7013)
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3
US-09-177-249-6
Query Match
Best Local Similarity 7.0%; Score 16; DB 3; Length 7015;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 ATTTCTATTCTTGC 126
Db 1319 ATTTCTATTCTTGC 1314
RESULT 18
US-09-922-445-1
Sequence 1, Application US/09922445
Patent No. 6528268
GENERAL INFORMATION:
APPLICANT: Anderson, Maria K.
APPLICANT: Berglund, Lars G. T.
APPLICANT: Reneland, Rikard H.
APPLICANT: Adam, Gail I. R.
TITLE OF INVENTION: REAGENTS AND METHODS FOR DETECTION OF HEART FAILURE
FILE REFERENCE: G6126US
CURRENT APPLICATION NUMBER: US/09/922,445
CURRENT FILING DATE: 2001-08-03
NUMBER OF SEQ ID NOS: 51
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 38653
TYPE: DNA
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)..(26156)
OTHER INFORMATION:
NAME/KEY: misc feature
LOCATION: (24801)..(24801)
OTHER INFORMATION: nucleotide 24801 is a single nucleotide polymorphism which can be
NAME/KEY: A or G
LOCATION: (24941)..(24941)
OTHER INFORMATION: nucleotide 24941 is a single nucleotide polymorphism which can be
OTHER INFORMATION: T or C
NAME/KEY: exon
LOCATION: (26157)..(26252)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (26253)..(26401)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (26402)..(26543)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (26544)..(27024)
```

```
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (327025)..(327178)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (327179)..(30519)
OTHER INFORMATION:
NAME/KEY: misc feature
LOCATION: (327645)..(327645)
OTHER INFORMATION: nucleotide 27645 is a single nucleotide polymorphism which can be
OTHER INFORMATION: C or G
NAME/KEY: exon
LOCATION: (30520)..(30681)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (30682)..(30894)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (30895)..(31027)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (31028)..(31747)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (31748)..(31841)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (31842)..(32400)
OTHER INFORMATION:
NAME/KEY: misc feature
LOCATION: (32163)..(32163)
OTHER INFORMATION: nucleotide 32163 is a single nucleotide polymorphism which can be
OTHER INFORMATION: A or C
NAME/KEY: exon
LOCATION: (32401)..(32528)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (32529)..(33414)
OTHER INFORMATION:
NAME/KEY: misc feature
LOCATION: (32614)..(32614)
OTHER INFORMATION: nucleotide 32614 is a single nucleotide polymorphism which can be
OTHER INFORMATION: A or G
NAME/KEY: exon
LOCATION: (33415)..(33597)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (33598)..(34314)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (34315)..(34588)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (34589)..(36404)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (36405)..(36523)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (36524)..(38341)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (38342)..(38653)
OTHER INFORMATION:
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Genbank/AC004923
DATABASE ENTRY DATE: 1999-12-21
RELEVANT RESIDUES: (1)..(38653)
US-09-922-445-1
```

```
QY 40 TTTTGATGGCCTTAAG 55
Db 19091 TTTTGATGGCCTTAAG 19106

RESULT 19
US-09-539-333D-1/c
Sequence 1, Application US/09539333D
Patent No. 6476208
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Bihain, Bernard
APPLICANT: Bastoux, Laurent
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
FILE REFERENCE: GENSET.047AUS
CURRENT FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: US/09/539,333D
PRIOR FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: US 60/126,903
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/131,971
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/132,065
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/143,928
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 60/145,915
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: US 60/146,453
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/146,452
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/162,288
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: US 09/416,384
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 31..1107
OTHER INFORMATION: 5'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 1108..1289
OTHER INFORMATION: exon A g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 14877..14920
OTHER INFORMATION: exon B g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 18778..18862
OTHER INFORMATION: exon Bbis g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 25593..25740
OTHER INFORMATION: exon C g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29388..29502
OTHER INFORMATION: exon D g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29967..30282
OTHER INFORMATION: exon E g35018 gene
```

```
Query Match 7.0%; Score 16; DB 4; Length 38653;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


FEATURE:
NAME/KEY: exon
LOCATION: 64666..64812
OTHER INFORMATION: exon F g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 65505..65853
OTHER INFORMATION: exon G g35018 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 65854..67854
OTHER INFORMATION: 3' regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 94124..94964
OTHER INFORMATION: exon g35017
FEATURE:
NAME/KEY: exon
LOCATION: 201188..201234
OTHER INFORMATION: exon S g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216836..216915
OTHER INFORMATION: exon V g35030 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 213818..215818
OTHER INFORMATION: 3' regulatory region g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215941
OTHER INFORMATION: exon R complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215975
OTHER INFORMATION: exon R complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..216952
OTHER INFORMATION: exon Qb1s complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..217061
OTHER INFORMATION: exon Q complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 217027..217061
OTHER INFORMATION: exon Q1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 229647..229742
OTHER INFORMATION: exon X complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 230408..230721
OTHER INFORMATION: exon P complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231272..231412
OTHER INFORMATION: exon Ob1s complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231787..231880
OTHER INFORMATION: exon O2 complement g34872 gene
FEATURE:

NAME/KEY: exon
LOCATION: 231870..231879
OTHER INFORMATION: exon O1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 234174..234321
OTHER INFORMATION: exon O complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 237406..237428
OTHER INFORMATION: exon Nb1s complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239807
OTHER INFORMATION: exon N2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239853
OTHER INFORMATION: exon N complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240569
OTHER INFORMATION: exon M117 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240596
OTHER INFORMATION: exon M1090 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240617
OTHER INFORMATION: exon M1069 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240644
OTHER INFORMATION: exon M52 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240824
OTHER INFORMATION: exon M862 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240994
OTHER INFORMATION: exon M692 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..241685
OTHER INFORMATION: exon M1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240800..240993
OTHER INFORMATION: exon M51 complement g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 241686..243685
OTHER INFORMATION: 5' regulatory region g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 290652..292652
OTHER INFORMATION: 3' regulatory region g34665 gene
FEATURE:
NAME/KEY: exon
LOCATION: 292653..292841

Query Match 7.0%; Score 16; DB 4; Length 319608;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 114 TTGATTCTTGCTC 129
|||||
Db 135178 TTGATTCTTGCTC 135163

RESULT 20

```
US-09-679-409-1/c
; Sequence 1, Application US/09679409
; Patent No. 655316
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Essieux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US15.CIP
; CURRENT APPLICATION NUMBER: US/09/679,409
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 09/416,384
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 60/168,088
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 199122..201122
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 201123..201234
; OTHER INFORMATION: exon S
; NAME/KEY: exon
; LOCATION: 201123..201560
; OTHER INFORMATION: exon S2
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U
; NAME/KEY: exon
; LOCATION: 216836..216994
; OTHER INFORMATION: exon V
; NAME/KEY: exon
; LOCATION: 216836..217077
; OTHER INFORMATION: exon V2
; NAME/KEY: exon
; LOCATION: 217671..217764
; OTHER INFORMATION: exon V1
; NAME/KEY: exon
; LOCATION: 227655..227736
; OTHER INFORMATION: exon V4
; NAME/KEY: exon
; LOCATION: 238715..238919
; OTHER INFORMATION: exon V3
; NAME/KEY: exon
; LOCATION: 240440..240673
; OTHER INFORMATION: exon W
; NAME/KEY: exon
; LOCATION: 240440..241153
; OTHER INFORMATION: exon W2
; NAME/KEY: exon
; LOCATION: 241072..241291
; OTHER INFORMATION: exon X
; NAME/KEY: exon
; LOCATION: 244353..244561
; OTHER INFORMATION: exon Y
; NAME/KEY: exon
; LOCATION: 246273..247802
; OTHER INFORMATION: exon Z
; NAME/KEY: misc.feature
; LOCATION: 247803..249803

; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 8316
; OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 21672
; OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 65485
; OTHER INFORMATION: 8-128-33 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 95396
; OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 107281
; OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 160640
; OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 160876
; OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 168974
; OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 170810
; OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 173358
; OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 189957
; OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 197163
; OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 200778
; OTHER INFORMATION: 8-303-235 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 202651
; OTHER INFORMATION: 8-300-221 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 202679
; OTHER INFORMATION: 8-300-193 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 203378
; OTHER INFORMATION: 8-299-128 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 204138
; OTHER INFORMATION: 8-296-213 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 204605
; OTHER INFORMATION: 8-252-190 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 204934
; OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 205206
; OTHER INFORMATION: 8-295-248 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 205329
; OTHER INFORMATION: 8-295-125 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 206064
; OTHER INFORMATION: 8-293-130 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 206545
; OTHER INFORMATION: 8-292-198 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 207313
; OTHER INFORMATION: 8-251-322 : polymorphic base A or G
```

```
/ NAME/KEY: allele
/ LOCATION: 208285 : polymorphic base A or G
/ OTHER INFORMATION: 8-289-322 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 208960 : polymorphic base C or T
/ OTHER INFORMATION: 8-287-249 : polymorphic base C or T
/ NAME/KEY: allele
/ LOCATION: 209123 : polymorphic base A or T
/ OTHER INFORMATION: 8-287-86 : polymorphic base A or T
/ NAME/KEY: allele
/ LOCATION: 209631 : polymorphic base A or G
/ OTHER INFORMATION: 8-285-319 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 210361 : polymorphic base G or C
/ OTHER INFORMATION: 8-283-278 : polymorphic base G or C
/ NAME/KEY: allele
/ LOCATION: 210463 : polymorphic base A or G
/ OTHER INFORMATION: 8-283-176 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 210486 : polymorphic base G or C
/ OTHER INFORMATION: 8-283-153 : polymorphic base G or C
/ NAME/KEY: allele
/ LOCATION: 210583 : polymorphic base C or T
/ OTHER INFORMATION: 8-283-56 : polymorphic base C or T
/ NAME/KEY: allele
/ LOCATION: 210879 : polymorphic base G or C
/ OTHER INFORMATION: 8-282-345 : polymorphic base G or C
/ NAME/KEY: allele
/ LOCATION: 210964 : polymorphic base G or T
/ OTHER INFORMATION: 8-282-260 : polymorphic base G or T
/ NAME/KEY: allele
/ LOCATION: 210979 : polymorphic base A or C
/ OTHER INFORMATION: 8-282-245 : polymorphic base A or C
/ NAME/KEY: allele
/ LOCATION: 211050 : polymorphic base A or T
/ OTHER INFORMATION: 8-282-174 : variable motif AAAG or GAAGGAAGGAAGGAAGA
/ NAME/KEY: allele
/ LOCATION: 211132 : polymorphic base A or T
/ OTHER INFORMATION: 8-282-92 : polymorphic base A or T
/ NAME/KEY: allele
/ LOCATION: 211247 : polymorphic base A or G
/ OTHER INFORMATION: 8-281-367 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 211315 : polymorphic base A or G
/ OTHER INFORMATION: 8-281-299 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 211366 : polymorphic base G or C
/ OTHER INFORMATION: 8-281-248 : polymorphic base G or C
/ NAME/KEY: allele
/ LOCATION: 212520 : polymorphic base A or C
/ OTHER INFORMATION: 8-279-197 : polymorphic base A or C
/ NAME/KEY: allele
/ LOCATION: 212821 : polymorphic base C or T
/ OTHER INFORMATION: 8-278-289 : polymorphic base C or T
/ NAME/KEY: allele
```

Query Match 7.0%; Score 16; DB 4; Length 319608;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 TTCATTCTTGCTC 129
Db 135178 TTCTATTCTTGCTC 135163

RESULT 21
US-09-221-017B-254
Sequence 254, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120

```
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORRISON & FOERSTER
/ STREET: 755 PAGE MILL ROAD
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304-1018
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows
/ SOFTWARE: FASTSEQ for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/221,017B
/ FILING DATE: 23-DEC-1998
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PP1182
/ FILING DATE: 31-DEC-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PP1546
/ FILING DATE: 30-JAN-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PP2911
/ FILING DATE: 09-APR-1998
/ APPLICATION DATA:
/ APPLICATION NUMBER: PCT/AU98/01023
/ FILING DATE: 10-DEC-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Montoy, Gladys H
/ REGISTRATION NUMBER: 32,430
/ REFERENCE/DOCKET NUMBER: 27340-20021.00
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-813-5600
/ TELEFAX: 650-494-0792
/ TELEX: 706141
/ INFORMATION FOR SEQ ID NO: 254:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 920 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHEetical: NO
/ ANTI-SENSE: UNKNOWN
/ ORIGINAL SOURCE:
/ ORGANISM: PORYPHYROMONAS GINGIVALIS
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1...920
/ US-09-221-017B-254
```

Query Match 6.6%; Score 15; DB 4; Length 920;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 CGAGTCTGATTT 179
Db 672 CGAGTCTGATTT 686

RESULT 22
US-09-252-991A-51/C
Sequence 51, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 51
LENGTH: 957
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (102)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-51

Query Match
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 TTCTATTCTTGCC 127
155 TTCTATTCTTGCC 141

Db

RESULT 23
US-09-107-532A-1108
Sequence 1108, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arimello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1108:
SEQUENCE CHARACTERISTICS:
LENGTH: 975 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...975

SEQUENCE DESCRIPTION: SEQ ID NO: 1108:
US-09-107-532A-1108

Query Match
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 CAGAAATTTTGATG 47
665 CAGAAATTTTGATG 679

Db

RESULT 24
US-09-784-508-16
Sequence 16, Application US/09784508
Patent No. 6548742
GENERAL INFORMATION:
APPLICANT: Martin, Robert R.
APPLICANT: Mathews, Helena
APPLICANT: Keller, Karen
APPLICANT: Kellogg, Jill A.
APPLICANT: Wagner, Ry
TITLE OF INVENTION: Development of Resistance to Raspberry
FILE REFERENCE: 4257-0024.31
CURRENT APPLICATION NUMBER: US/09/784,508
CURRENT FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: US 09/737,719
PRIOR FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: US 60/171,018
PRIOR FILING DATE: 1999-01-25
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 1059
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: modified movement protein coding sequence RBDV4
US-09-784-508-16

Query Match
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 AAGTCTCTACTCG 67
12 AAGTCTCTACTCG 26

Db

RESULT 25
US-09-784-508-3
Sequence 3, Application US/09784508
Patent No. 6548742
GENERAL INFORMATION:
APPLICANT: Martin, Robert R.
APPLICANT: Mathews, Helena
APPLICANT: Keller, Karen
APPLICANT: Kellogg, Jill A.
APPLICANT: Wagner, Ry
TITLE OF INVENTION: Development of Resistance to Raspberry
FILE REFERENCE: 4257-0024.31
CURRENT APPLICATION NUMBER: US/09/784,508
CURRENT FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: US 09/737,719
PRIOR FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: US 60/171,018
PRIOR FILING DATE: 1999-01-25
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1077

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence encoding RBDV movement protein
US-09-784-508-3

```

Query Match	6.6%	Score 15;	DB 4;	Length 1077;
Best Local Similarity	100.0%	Pred. No. 84;		
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	53	AAGTTCCTCTACTCG	67
Db	12	AAGTTCCTCTACTCG	26

```

RESULT 26
US-09-328-352-2461
; Sequence 2461, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2461
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-2461

```

Query Match	6.6%;	Score 15;	DB 4;	Length 1491;
Best Local Similarity	100.0%;	Pred. No. 85;		
Matches	15;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

QY	119	TTTCTTGCCCTCGTAT	133
Db	90	TTTCTTGCCCTCGTAT	104

US-09-27
US-09-697-367-7/c
Sequence 7, Application US/09697367
Patent No. 632015
GENERAL INFORMATION:
APPLICANT: Orozco Jr., Emil M.
APPLICANT: Calmi, Perry G.
APPLICANT: Weng, Zude
APPLICANT: Tarczynski, Mitchell
TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: B6116 US NA
CURRENT APPLICATION NUMBER: US/09/697,367
CURRENT FILING DATE: 2000-10-26
PRIOR APPLICATION NUMBER: 66/084,529
PRIOR FILING DATE: 1998-MAY-07
PRIOR APPLICATION NUMBER: PCT/US99/09865
PRIOR FILING DATE: 1999-MAY-06
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 7
LENGTH: 1611
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: unsure
LOCATION: (14)
NAME/KEY: unsure
LOCATION: (1424)
NAME/KEY: unsure
LOCATION: (1469)
NAME/KEY: unsure

```

; LOCATION: (1505)
; NAME/KEY: unsure
; LOCATION: (1515)
; NAME/KEY: unsure
; LOCATION: (1560)
;
US-09-697-367-7

```

Query Match 6.6%; Score 15; DB 4; Length 1611;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	114	TTCTATTCTTGCCCT	128
Db	547	TTCTATTCTTGCCCT	533

```

RESULT 28
US-08-936-165A-224
: Sequence 224, Application US/08936165A
: Patent No. 6348582
:
: GENERAL INFORMATION:
:
: APPLICANT: Black, Michael
: APPLICANT: Burnham, Martin
: APPLICANT: Hodgson, John
: APPLICANT: Knowles, David
: APPLICANT: Lonetto, Michael
: APPLICANT: Nicholas, Richard
: APPLICANT: Pratt, Julie
: APPLICANT: Reichard, Richard
: APPLICANT: Rosenberg, Martin
: APPLICANT: Ward, Judith
:
: TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides
: TITLE OF INVENTION: Polypeptides and their uses
:
: NUMBER OF SEQUENCES: 534

```

ADDRESS: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,165A
FILING DATE: 24-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027,032
FILING DATE: 24-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm1, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090

```

; INFORMATION FOR SEQ ID NO: 224
;
; SEQUENCE CHARACTERISTICS:
;
;   LENGTH: 1886 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
; MOLECULE TYPE: Genomic DNA
;
; US-08-936-165A-224

```

Query Match	6.6%;	Score 15;	DB 4;	Length 1886;
Best Local Similarity	100.0%;	Pred. No. 86;		
Matches	15;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

Qy 111 ATTTCTATTCTTG 125
|||||
Db 205 ATTTCTATTCTTG 219

RESULT 29

US-08-933-821-16/c
Sequence 16, Application US/08933821
Patent No. 5972338

GENERAL INFORMATION:

APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Tie ligands
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,821

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1130
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 2042 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-933-821-16

Query Match 6.6%; Score 15; DB 2; Length 2042;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 GATTTCTATTCTT 124
|||||
Db 640 GATTTCTATTCTT 626

RESULT 30

US-08-934-494-5/c
Sequence 5, Application US/08934494
Patent No. 6030831

GENERAL INFORMATION:

APPLICANT: Gurney, Austin
TITLE OF INVENTION: Tie ligands
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,494
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1078
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 2042 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-934-494-5

Query Match 6.6%; Score 15; DB 3; Length 2042;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 GATTTCTATTCTT 124
|||||
Db 640 GATTTCTATTCTT 626

RESULT 31

US-08-960-507-16/c
Sequence 16, Application US/08960507
Patent No. 6057435

GENERAL INFORMATION:

APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Tie ligands
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,507

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1130p1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 2042 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-960-507-16

Query Match 6.6%; Score 15; DB 3; Length 2042;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCATTCTT 124
|||
Db 640 GATTTCATTCTT 626

RESULT 32

US-09-143-068-5/c

Sequence 5, Application US/09143068

Patent No. 6074873

GENERAL INFORMATION:

APPLICANT: Fong, Sherman

APPLICANT: Ferrara, Napoleone

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth

TITLE OF INVENTION: Tie Ligand Homologues

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/143,068

FILING DATE: 28-Aug-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: P1078P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2042 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-09-143-068-5

Query Match 6.6%; Score 15; DB 3; Length 2042;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCATTCTT 124

|||

Db 640 GATTTCATTCTT 626

RESULT 33

US-09-143-707-5/c

Sequence 5, Application US/09143707

Patent No. 6348350

GENERAL INFORMATION:

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

TITLE OF INVENTION: Tie Ligands

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/143,707

FILING DATE: 28-Aug-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: P1078P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2042 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-09-143-707-5

Query Match 6.6%; Score 15; DB 4; Length 2042;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCATTCTT 124

|||

Db 640 GATTTCATTCTT 626

RESULT 34

US-09-202-089-5/c

Sequence 5, Application US/09202089

Patent No. 6348351

GENERAL INFORMATION:

APPLICANT: Genentech Inc.

APPLICANT: Fong, Sherman

APPLICANT: Ferrara, Napoleone

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth

APPLICANT: Williams, Mickey

TITLE OF INVENTION: Tie Ligand Homologues

FILE REFERENCE: P1078P2(e)

CURRENT APPLICATION NUMBER: US/09/202,089

CURRENT FILING DATE: 1998-12-08

PRIOR APPLICATION NUMBER: PCT/US98/19094

PRIOR FILING DATE: 1998-09-14

PRIOR APPLICATION NUMBER: US 08/934,494

PRIOR FILING DATE: 1997-09-19

NUMBER OF SEQ ID NOS: 17

SEQ ID NO 5

LENGTH: 2042

TYPE: DNA

ORGANISM: Homo sapiens

US-09-202-089-5

Query Match 6.6%; Score 15; DB 4; Length 2042;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCATTCTT 124

|||

Db 640 GATTTCATTCTT 626

```
RESULT 35
US-09-136-828-16/c
; Sequence 16, Application US/09136828
; Patent No. 6350450
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,828
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130R1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-09-136-828-16

Query Match          6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      110 GATTTCTATTCTT 124
        |||||||
        640 GATTTCTATTCTT 626

Db

RESULT 36
US-09-332-928A-16/c
; Sequence 16, Application US/09332928A
; Patent No. 6368853
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,928A
```

```
; FILING DATE: 14-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/933,821
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-332-928A-16

Query Match          6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      110 GATTTCTATTCTT 124
        |||||||
        640 GATTTCTATTCTT 626

Db

RESULT 37
US-09-511-133-5/c
; Sequence 5, Application US/09511133
; Patent No. 6372491
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; FILE REFERENCE: PI078R1c1
; CURRENT APPLICATION NUMBER: US/09/511,133
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 09/143,707
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 60/059,352
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 5
; LENGTH: 2042
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-511-133-5

Query Match          6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      110 GATTTCTATTCTT 124
        |||||||
        640 GATTTCTATTCTT 626

Db

RESULT 38
US-09-136-801-16/c
; Sequence 16, Application US/09136801
; Patent No. 6413770
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Hillan, Kenneth
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
; APPLICANT: Roy, Margaret
```



```

;
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Tumas, Daniel
; APPLICANT: Schwall, Ralph
; TITLE OF INVENTION: The Ligand Homologues
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPacIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,801
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
;
; US-09-136-801-16
;
Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
Db 640 GATTTCTATTCTT 626

RESULT 39
US-09-332-929-16/c
; Sequence 16, Application US/09332929
; Patent No. 6420542
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: The Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPacIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,929
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: 08/933,821
; FILING DATE:

```

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
;
; US-09-332-929-16
;
Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
Db 640 GATTTCTATTCTT 626

RESULT 40
US-09-690-169-5/c
; Sequence 5, Application US/09690169
; Patent No. 6426218
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: THE LIGANDS
; FILE REFERENCE: GENE 58A2D2
; CURRENT APPLICATION NUMBER: US/09/690,169
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 09/143,707
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 60/059,352
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2042
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; US-09-690-169-5
;
Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
Db 640 GATTTCTATTCTT 626

RESULT 41
US-09-511-631-5/c
; Sequence 5, Application US/09511631
; Patent No. 6455496
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: The Ligands
; FILE REFERENCE: P1078R1D1
; CURRENT APPLICATION NUMBER: US/09/511,631
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 09/143,707
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 60/059,352
; PRIOR FILING DATE: 1997-09-19

```

; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 5
; LENGTH: 2042
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-511-631-5

Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
|||||
Db 640 GATTTCTATTCTT 626

RESULT 42
US-09-333-075-16/c
; Sequence 16, Application US/09333075
; Patent No. 6492331
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/933,821
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9981

; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-09-333-075-16

Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
|||||
Db 640 GATTTCTATTCTT 626

RESULT 43
US-09-690-189-5/c
; Sequence 5, Application US/09690189
; Patent No. 6521234
; GENERAL INFORMATION:

; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; FILE REFERENCE: P1078RID3
; CURRENT APPLICATION NUMBER: US/09/690,189
; CURRENT FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: US 09/143,707
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 5
; LENGTH: 2042
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-690-189-5

Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
|||||
Db 640 GATTTCTATTCTT 626

RESULT 44
US-09-202-088A-16/c
; Sequence 16, Application US/09202088A
; Patent No. 6551822
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
; APPLICANT: Roy, Margaret
; APPLICANT: Feitara, Napoleone
; APPLICANT: Tumaes, Daniel
; APPLICANT: Schwall, Ralph
; TITLE OF INVENTION: TIE LIGAND HOMOLOGUES
; FILE REFERENCE: P1130P3US
; CURRENT APPLICATION NUMBER: US/09/202,088A
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: PCT/US98/19093
; PRIOR FILING DATE: 1998-09-14
; NUMBER OF SEQ ID NOS: 35
; SEQ ID NO 16
; LENGTH: 2042
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-202-088A-16

Query Match 6.6%; Score 15; DB 4; Length 2042;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTT 124
|||||
Db 640 GATTTCTATTCTT 626

RESULT 45
US-09-333-077-16/c
; Sequence 16, Application US/09333077
; Patent No. 6586397
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way

CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Winpatin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/333,077
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/933,821
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Dreyer, Ginger R.
 REGISTRATION NUMBER: 33,055
 REFERENCE/DOCKET NUMBER: P1130
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650/225-3216
 TELEFAX: 650/952-9881
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2042 base pairs
 TYPE: Nucleic Acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 US-09-333-077-16

Query Match 6.6%; Score 15; DB 4; Length 2042;
 Best Local Similarity 100.0%; Pred. NO. 86; Mismatches 0; Indels 0; Gaps 0;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 110 GATTTCTATTCTT 124
 |||||||||
 Db 640 GATTTCTATTCTT 626

Search completed: February 4, 2004, 16:08:09
 Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 14:32:00 ; Search time 1688 Seconds
(without alignments)
497.552 Million cell updates/sec

Title: US-09-675-650-1_COPY_27_254
Perfect score: 228
Sequence: 1 gtaagcgttataaagcac.....tttcctccacattcacag 228

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 2449703 seqs, 1841816367 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/PCrUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
 - 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
 - 15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
 - 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
 - 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
 - 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	7.9	476	11	US-09-918-995-12566
C 2	18	7.9	476	13	US-10-027-632-302898
C 3	18	7.9	476	14	US-10-027-632-302898
C 4	18	7.9	481	13	US-10-027-632-195312
C 5	18	7.9	481	14	US-10-027-632-195312
C 6	18	7.9	637	13	US-10-027-632-199346
C 7	18	7.9	637	14	US-10-027-632-199346
C 8	18	7.9	637	13	US-10-027-632-199346
C 9	18	7.9	637	14	US-10-027-632-199346
C 10	18	7.9	301692	12	US-10-027-632-199347
C 11	17	7.5	378	11	US-09-918-995-18845
C 12	17	7.5	447	12	US-10-242-535A-23845
C 13	17	7.5	450	12	US-10-242-535A-23921
C 14	17	7.5	581	13	US-10-027-632-310795
C 15	17	7.5	581	14	US-10-027-632-310795

16	17	7.5	601	10	US-09-818-656A-94	Sequence 94, Appl
17	17	7.5	601	10	US-09-818-656A-95	Sequence 95, Appl
18	17	7.5	724	13	US-10-027-632-149573	Sequence 149573,
C 19	17	7.5	724	13	US-10-027-632-149574	Sequence 149574,
C 20	17	7.5	724	14	US-10-027-632-149573	Sequence 149573,
C 21	17	7.5	724	14	US-10-027-632-149574	Sequence 149574,
C 22	17	7.5	927	9	US-09-815-242-7429	Sequence 7429, Ap
C 23	17	7.5	2279	13	US-10-191-997-1310	Sequence 130, Ap
C 24	17	7.5	2318	15	US-10-117-722-916	Sequence 916, Ap
C 25	17	7.5	2318	15	US-10-037-270-916	Sequence 916, Ap
C 26	17	7.5	6352	11	US-09-764-891-7122	Sequence 7122, Ap
C 27	17	7.5	18657	15	US-10-074-045-70	Sequence 70, Appl
C 28	17	7.5	28313	10	US-09-764-877-3194	Sequence 3194, Ap
C 29	17	7.5	28313	12	US-10-242-515-3194	Sequence 3194, Ap
C 30	17	7.5	29228	10	US-09-764-877-3198	Sequence 3198, Ap
C 31	17	7.5	29228	12	US-10-242-515-3198	Sequence 3198, Ap
C 32	17	7.5	40645	10	US-09-818-656A-3	Sequence 3, Appl1
C 33	17	7.5	40645	14	US-10-216-441-3	Sequence 3, Appl1
C 34	16	7.0	227	12	US-10-242-535A-29962	Sequence 29962, A
C 35	16	7.0	287	10	US-09-920-300A-1718	Sequence 1718, Ap
C 36	16	7.0	287	13	US-10-099-926-1718	Sequence 1718, Ap
C 37	16	7.0	287	14	US-10-033-528-1718	Sequence 1718, Ap
C 38	16	7.0	370	14	US-10-040-739-1242	Sequence 1242, Ap
C 39	16	7.0	422	10	US-09-796-692-3452	Sequence 3452, Ap
C 40	16	7.0	422	12	US-10-057-475B-3452	Sequence 3452, Ap
C 41	16	7.0	422	12	US-10-154-884B-3452	Sequence 3452, Ap
C 42	16	7.0	422	15	US-10-040-862-3452	Sequence 3452, Ap
C 43	16	7.0	427	10	US-09-796-692-2901	Sequence 2901, Ap
C 44	16	7.0	427	12	US-10-057-475B-2901	Sequence 2901, Ap
C 45	16	7.0	427	12	US-10-154-884B-2901	Sequence 2901, Ap

ALIGNMENTS

RESULT 1
US-09-918-995-12566/c
; Sequence 12566, Application US/09918995
; Publication No. US20030073623A1
GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918, 995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12566
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)..(476)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-12566

Query Match 7.9%; Score 18; DB 11; Length 476;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ATTTTGATGCGCTTAAAG 55
Db 110 ATTTTGATGCGCTTAAAG 93
RESULT 2
US-10-027-632-302898/c
; Sequence 302898, Application US/10027632
; Publication No. US20030204075A9

```

1  GENERAL INFORMATION:
2  APPLICANT: Wang, David G.
3  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
4  TITLE OF INVENTION: Polymorphisms in the Human Genome
5  FILE REFERENCE: 108827.129
6  CURRENT APPLICATION NUMBER: US/10/0277,632
7  CURRENT FILING DATE: 2002-04-30
8  PRIOR APPLICATION NUMBER: US 60/218,006
9  PRIOR FILING DATE: 2000-07-12
10 PRIOR APPLICATION NUMBER: US 60/198,676
11 PRIOR FILING DATE: 2000-04-20
12 PRIOR APPLICATION NUMBER: US 60/193,483
13 PRIOR FILING DATE: 2000-03-29
14 PRIOR APPLICATION NUMBER: US 60/185,218
15 PRIOR FILING DATE: 2000-02-24
16 PRIOR APPLICATION NUMBER: US 60/167,363
17 PRIOR FILING DATE: 1999-11-23
18 PRIOR APPLICATION NUMBER: US 60/156,358
19 PRIOR FILING DATE: 1999-09-28
20 PRIOR APPLICATION NUMBER: US 60/146,002
21 PRIOR FILING DATE: 1999-08-09
22 NUMBER OF SEQ ID NOS: 325720
23 SOFTWARE: FastSeq for Windows Version 4.0
24 SEQ ID NO 302898
25 LENGTH: 476
26 TYPE: DNA
27 ORGANISM: Human
28 US-10-027-632-302898

```

Query Match	7.9%	Score 18	DB 13	Length 476
Best Local Similarity	100.0%	Pred. No. 14		
Matches	18	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0
Qy	21	TCOAATTTCTACTCAGAAA	38	
db	120	TCOAATTTCTACTCAGAAA	103	

```

RESULT 3
US-10-027-632-302698/c
; Sequence 302698, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.139
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 302698
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-302698

```

Query Match	7.9%	Score 18	DB 14	Length 476
Best Local Similarity	100.0%	Pred. No. 14		
Matches 18; Conservative	0	Mismatches	0	Gaps 0

QY	21	TC	AT	TT	CT	ACT	CAG	AA	38
Db	120	TC	AT	TT	CT	ACT	CAG	AA	103

```

RESULT 4
US-10-027-632-195212
; Sequence 195212, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199, 676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193, 483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185, 218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167, 353
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156, 358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146, 002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: Pat-Seq for Windows Version 4.0
; SEQ ID NO 195212
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195212

```

	Query Match	7.9%	Score 18	DB 13	Length 481
	Similarity	100.0%	Pred. No. 14		
	Matches 18	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	21 TCAATTTCTACTCAGAA	38			
Db	346 TCAATTTCTACTCAGAA	363			

```

? RESULT 5
? US-10-027-632-195212
? Sequence 195212, Application US/10027632
? GENERAL INFORMATION:
? APPLICANT: Wang, David G.
? TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
? Polymorphisms in the Human Genome
? FILE REFERENCE: 108827.129
? CURRENT APPLICATION NUMBER: US/10/027,632
? CURRENT FILING DATE: 2002-04-30
? PRIOR APPLICATION NUMBER: US 60/218, 006
? PRIOR FILING DATE: 2000-07-12
? PRIOR APPLICATION NUMBER: US 60/198, 676
? PRIOR FILING DATE: 2000-04-20
? PRIOR APPLICATION NUMBER: US 60/193, 483
? PRIOR FILING DATE: 2000-03-29
? PRIOR APPLICATION NUMBER: US 60/165, 218
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: US 60/167, 363
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: US 60/156, 358
? PRIOR FILING DATE: 1999-09-28
? PRIOR APPLICATION NUMBER: US 60/146, 002
? PRIOR FILING DATE: 1999-08-09
? NUMBER OF SEQ ID NOS: 325720

```

SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195212
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-195212

Query Match 7.9%; Score 18; DB 14; Length 481;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 TCAATTTCTACTCGAAA 38
|||||
Db 346 TCAATTTCTACTCGAAA 363

RESULT 6
US-10-027-632-199346/c
; Sequence 199346, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199346
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-199346

Query Match 7.9%; Score 18; DB 13; Length 637;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATAAAGCACTCAATTCT 29
|||||
Db 465 ATAAAGCACTCAATTCT 448

RESULT 7
US-10-027-632-199347/c
; Sequence 199347, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199347
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-199347

Query Match 7.9%; Score 18; DB 13; Length 637;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATAAAGCACTCAATTCT 29
|||||
Db 465 ATAAAGCACTCAATTCT 448

RESULT 8
US-10-027-632-199346/c
; Sequence 199346, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199346
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-199346

Query Match 7.9%; Score 18; DB 14; Length 637;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATAAAGCACTCAATTCT 29
|||||
Db 465 ATAAAGCACTCAATTCT 448

RESULT 9
US-10-027-632-199347/c
; Sequence 199347, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027.632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 199347
;; LENGTH: 637
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-199347

Query Match 7.9%; Score 18; DB 14; Length 637;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Indels 0; Gaps 0;

QY 12 ATAAAGCACTCAATTCT 29
Db 465 ATAAAGCACTCAATTCT 448

RESULT 10
US-10-428-487-11/c
;; Sequence 11, Application US/10428487
;; Publication No. US2004006780A1
;; GENERAL INFORMATION:
;; APPLICANT: RASTELTI, LUCA K.
;; APPLICANT: GERBER, HANS-PETER
;; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
;; FILE REFERENCE: 09800080-0103
;; CURRENT APPLICATION NUMBER: US/10/428.487
;; CURRENT FILING DATE: 2003-05-02
;; PRIOR APPLICATION NUMBER: 09/815,153
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,201
;; PRIOR FILING DATE: 2000-03-22
;; NUMBER OF SEQ ID NOS: 84
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 11
;; LENGTH: 301692
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-428-487-11

Query Match 7.9%; Score 18; DB 12; Length 301692;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ATAAAGCACTCAATTCT 29
Db 147492 ATAAAGCACTCAATTCT 147475

RESULT 11
US-09-918-18845/c
;; Sequence 18845, Application US/09918995
;; Publication No. US20030073623A1
;; GENERAL INFORMATION:

;; APPLICANT: Hyseq, Inc.
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
;; FILE REFERENCE: 20411-756
;; CURRENT APPLICATION NUMBER: US/09/918.995
;; CURRENT FILING DATE: 2001-07-30
;; PRIOR APPLICATION NUMBER: US/09/235,076
;; PRIOR FILING DATE: 1999-01-20
;; NUMBER OF SEQ ID NOS: 38054
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 18845
;; LENGTH: 378
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(378)
;; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-18845

Query Match 7.5%; Score 17; DB 11; Length 378;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Indels 0; Gaps 0;

QY 27 TCTACTCAGAAATTTT 43
Db 178 TCTACTCAGAAATTTT 162

RESULT 12
US-10-242-535A-23845
;; Sequence 23845, Application US/10242535A
;; Publication No. US20040013663A1
;; GENERAL INFORMATION:
;; APPLICANT: Llew, C.C.
;; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
;; FILE REFERENCE: 4231/2005
;; CURRENT APPLICATION NUMBER: US/10/242.535A
;; CURRENT FILING DATE: 2002-09-12
;; PRIOR APPLICATION NUMBER: US 10/085,783
;; PRIOR FILING DATE: 2002-02-28
;; PRIOR APPLICATION NUMBER: US 60/305,340
;; PRIOR FILING DATE: 2001-07-13
;; PRIOR APPLICATION NUMBER: US 60/275,017
;; PRIOR FILING DATE: 2001-03-12
;; PRIOR APPLICATION NUMBER: US 60/271,955
;; PRIOR FILING DATE: 2001-02-28
;; NUMBER OF SEQ ID NOS: 58994
;; SOFTWARE: PatentIn Version 3.2
;; SEQ ID NO 23845
;; LENGTH: 447
;; TYPE: DNA
;; ORGANISM: Human
US-10-242-535A-23845

Query Match 7.5%; Score 17; DB 12; Length 447;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TCTACTCAGAAATTTT 43
Db 220 TCTACTCAGAAATTTT 236

RESULT 13
US-10-242-535A-25921
;; Sequence 25921, Application US/10242535A
;; Publication No. US20040013663A1
;; GENERAL INFORMATION:
;; APPLICANT: ChondroGene Inc.
;; APPLICANT: Llew, C.C.
;; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis


```
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25921
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (197)..(197)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (199)..(199)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (267)..(267)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (283)..(283)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (291)..(291)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-242-535A-25921

Query Match
Best Local Similarity 7.5%; Score 17; DB 12; Length 450;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 TCTACTCAGAAATTTT 43
Db 222 TCTACTCAGAAATTTT 238

RESULT 14
US-10-027-632-310795/c
; Sequence 310795, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 310795
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-310795

Query Match
Best Local Similarity 7.5%; Score 17; DB 13; Length 581;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TTTATAAGCCTCAAT 25
Db 482 TTTATAAGCCTCAAT 466

RESULT 15
US-10-027-632-310795/c
; Sequence 310795, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 310795
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-310795

Query Match
Best Local Similarity 7.5%; Score 17; DB 14; Length 581;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TTTATAAGCCTCAAT 25
Db 482 TTTATAAGCCTCAAT 466

RESULT 16
US-09-818-656A-94
; Sequence 94, Application US/09818656A
; Patent No. US20020142381A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CU001191
; CURRENT APPLICATION NUMBER: US/09/818,656A
; CURRENT FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94
```

LENGTH: 601
TYPE: DNA
ORGANISM: Homo sapiens
US-09-618-656A-94

Query Match 7.5%; Score 17; DB 10; Length 601;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 TTTCTATTCTTGCTC 129
|||||
Db 94 TTTCTATTCTTGCTC 110

RESULT 17
US-09-618-656A-95
Sequence 95, Application US/09818656A
Patent No. US20020142381A1
GENERAL INFORMATION:
APPLICANT: GONG, Fangcheng et al.
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001191
CURRENT APPLICATION NUMBER: US/09/818, 656A
CURRENT FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 103
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 95
LENGTH: 601
TYPE: DNA
ORGANISM: Homo sapiens
US-09-618-656A-95

Query Match 7.5%; Score 17; DB 10; Length 601;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 TTTCTATTCTTGCTC 129
|||||
Db 39 TTTCTATTCTTGCTC 55

RESULT 18
US-10-027-632-149573/C
Sequence 149573, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827,129
CURRENT APPLICATION NUMBER: US/10/027, 632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 149573
LENGTH: 724
TYPE: DNA

ORGANISM: Human
US-10-027-632-149573

Query Match 7.5%; Score 17; DB 13; Length 724;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ACTCAATTTCTACTCAG 35
|||||
Db 17 ACTCAATTTCTACTCAG 1

RESULT 19
US-10-027-632-149574/C
Sequence 149574, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827,129
CURRENT APPLICATION NUMBER: US/10/027, 632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 149574
LENGTH: 724
TYPE: DNA
ORGANISM: Human
US-10-027-632-149574

Query Match 7.5%; Score 17; DB 13; Length 724;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ACTCAATTTCTACTCAG 35
|||||
Db 17 ACTCAATTTCTACTCAG 1

RESULT 20
US-10-027-632-149573/C
Sequence 149573, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827,129
CURRENT APPLICATION NUMBER: US/10/027, 632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363

```

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149573
; LENGTH: 724
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-149573

Query Match
Best Local Similarity 100.0%; Score 17; DB 14; Length 724;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ACTCAATTTCTACTCAG 35
Db 17 ACTCAATTTCTACTCAG 1

RESULT 21
US-10-027-632-149574/c
; Sequence 149574, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149574
; LENGTH: 724
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-149574

Query Match
Best Local Similarity 100.0%; Score 17; DB 14; Length 724;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ACTCAATTTCTACTCAG 35
Db 17 ACTCAATTTCTACTCAG 1

RESULT 22
US-09-815-242-7429/c
; Sequence 7429, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
```

```

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7429
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(927)
US-09-815-242-7429

Query Match
Best Local Similarity 100.0%; Score 17; DB 9; Length 927;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GATTTCTATTCTTCTGC 126
Db 19 GATTTCTATTCTTCTGC 3

RESULT 23
US-10-191-997-130
; Sequence 130, Application US/10191997
; Publication No. US20030207834A1
; GENERAL INFORMATION:
; APPLICANT: Oligos Etc., Inc.
; APPLICANT: DALE, Roderic M. K.
; APPLICANT: ARROW, Amy
; APPLICANT: THOMPSON, Terry
; TITLE OF INVENTION: Oligonucleotide-Containing Pharmacological Compositions And Their
; FILE REFERENCE: 54800-5019
; CURRENT APPLICATION NUMBER: US/10/191,997
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/303,820
; PRIOR FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
; LENGTH: 2279
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: NM_001752
US-10-191-997-130

Query Match
Best Local Similarity 100.0%; Score 17; DB 13; Length 2279;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TCTACTCAGAAATTTT 43
```

Db 2044 TCTACTCAGAAATTTT 2060

RESULT 24

US-10-117-722-916
; Sequence 916, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; TITLE OF INVENTION: PolyPeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 916
; LENGTH: 2318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (103)..(1686)
US-10-117-722-916

Query Match 7.5%; Score 17; DB 13; Length 2318;

Best Local Similarity 100.0%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TCTACTCAGAAATTTT 43

Db 2078 TCTACTCAGAAATTTT 2094

RESULT 25

US-10-037-270-916
; Sequence 916, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Weinman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhwei
; APPLICANT: Tillinghaast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: PolyPeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 916
; LENGTH: 2318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (103)..(1686)
US-10-037-270-916

Query Match 7.5%; Score 17; DB 15; Length 2318;

Best Local Similarity 100.0%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 TCTACTCAGAAATTTT 43

Db 2078 TCTACTCAGAAATTTT 2094

RESULT 26

US-09-764-891-7122
; Sequence 7122, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7122
; LENGTH: 6352
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7122

Query Match 7.5%; Score 17; DB 11; Length 6352;

Best Local Similarity 100.0%; Pred. No. 59;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ATTTTGATGCGCTTAA 54

Db 924 ATTTTGATGCGCTTAA 940

RESULT 27

US-10-074-045-70
; Sequence 70, Application US/10074045
; Publication No. US20030092102A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12CI
; CURRENT APPLICATION NUMBER: US/10/074,045
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 18657
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-074-045-70

Query Match 7.5%; Score 17; DB 15; Length 18657;

Best Local Similarity 100.0%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 TTTTCATTTCTGCT 128
|||||
Db 6324 TTTTCATTTCTGCT 6340

RESULT 28

US-09-764-877-3194/c
; Sequence 3194, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3194
; LENGTH: 28313
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3194

Query Match 7.5%; Score 17; DB 10; Length 28313;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 CCTTCTACTACTGTC 91
|||||
Db 6228 CCTTCTACTACTGTC 6212

RESULT 29

US-10-242-515-3194/c
; Sequence 3194, Application US/10242515
; Publication No. US20040009488A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005C1
; CURRENT APPLICATION NUMBER: US/10/242,515
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3194
; LENGTH: 28313
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-242-515-3194

Query Match 7.5%; Score 17; DB 12; Length 28313;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 CCTTCTACTACTGTC 91
|||||
Db 6228 CCTTCTACTACTGTC 6212

RESULT 30

US-09-764-877-3198/c
; Sequence 3198, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3198
; LENGTH: 29228
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3198

Query Match 7.5%; Score 17; DB 10; Length 29228;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 CCTTCTACTACTGTC 91
|||||
Db 3774 CCTTCTACTACTGTC 3758

RESULT 31

US-10-242-515-3198/c
; Sequence 3198, Application US/10242515
; Publication No. US20040009488A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005C1
; CURRENT APPLICATION NUMBER: US/10/242,515
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3198
; LENGTH: 29228
; TYPE: DNA

ORGANISM: Homo sapiens
US-10-242-515-3198

Query Match 7.5%; Score 17; DB 12; Length 29228;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 CCTCTACTGACGTGC 91
Db 3774 CCTCTACTGACGTGC 3758

RESULT 32
US-09-818-656A-3
; Sequence 3, Application US/09818656A
; Patent No. US20020142381A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: C1001191
; CURRENT APPLICATION NUMBER: US/09/818, 656A
; CURRENT FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 40645
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-818-656A-3

Query Match 7.5%; Score 17; DB 10; Length 40645;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 TTCTATTCTTGCTC 129
Db 36001 TTCTATTCTTGCTC 36017

RESULT 33
US-10-216-441-3
; Sequence 3, Application US/10216441
; Publication No. US20020192762A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: C1001191CON
; CURRENT APPLICATION NUMBER: US/10/216, 441
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: 09/818, 656
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 40645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-216-441-3

Query Match 7.5%; Score 17; DB 14; Length 40645;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 TTCTATTCTTGCTC 129
Db 36001 TTCTATTCTTGCTC 36017

RESULT 34

US-10-242-535A-29962/c
; Sequence 29962, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Lew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242, 535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29962
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29962

Query Match 7.0%; Score 16; DB 12; Length 227;
Best Local Similarity 100.0%; Pred. No. 17e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TTTATTAAGCAGCTCAA 24
Db 37 TTTATTAAGCAGCTCAA 22

RESULT 35
US-09-920-300A-1718/c
; Sequence 1718, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Mesgher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920, 300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1718
; LENGTH: 287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 29, 215, 242, 257
; OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-1718

Query Match 7.0%; Score 16; DB 10; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 CTCACCTCGTCCCTC 204
Db 134 CTCACCTCGTCCCTC 119

RESULT 36
US-10-099-926-1718/c
; Sequence 1718, Application US/10099926

Publication No. US20030166064A1
GENERAL INFORMATION:
APPLICANT: King, Gordon E.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
APPLICANT: Jiang, Yugu
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.547C2
CURRENT APPLICATION NUMBER: US/10/099,926
CURRENT FILING DATE: 2002-03-17
NUMBER OF SEQ ID NOS: 1982
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1718
LENGTH: 287
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 29, 215, 242, 257
OTHER INFORMATION: n = A,T,C or G
US-10-099-926-1718

Query Match 7.0%; Score 16; DB 13; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTCACCTCCGTCCTC 204
|||||
DB 134 CTCACCTCCGTCCTC 119

RESULT 37
US-10-033-528-1718/C
Sequence 1718, Application US/10033528
Publication 1718, US20020131971A1
GENERAL INFORMATION:
APPLICANT: King, Gordon E.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.547C1
CURRENT APPLICATION NUMBER: US/10/033,528
CURRENT FILING DATE: 2001-12-26
NUMBER OF SEQ ID NOS: 1896
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1718
LENGTH: 287
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 29, 215, 242, 257
OTHER INFORMATION: n = A,T,C or G
US-10-033-528-1718

Query Match 7.0%; Score 16; DB 14; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CTCACCTCCGTCCTC 204
|||||
DB 134 CTCACCTCCGTCCTC 119

RESULT 38
US-10-040-739-1242/C
Sequence 1242, Application US/10040739
Publication No. US20020173635A1
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
McCoy, John
LaValle, Edward
Racie, Lisa
Merberg, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1242:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1242:
US-10-040-739-1242
Query Match 7.0%; Score 16; DB 14; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TTTATAAGCACTCA 24
|||||
DB 203 TTTATAAGCACTCA 188

RESULT 39
US-09-796-692-3452/C
Sequence 3452, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303

```

; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/222,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/222,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3452
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (151)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (165)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (340)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (347)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (356)
; OTHER INFORMATION: n=A,T,C or G
; US-09-796-692-3452

Query Match      7.0%; Score 16; DB 10; Length 422;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      154 ATTATCTCTCAGGAG 169
Db      291 ATTATCTCTCAGGAG 276

RESULT 40
US-10-057-475B-3452/C
; Sequence 3452, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
```

```

; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3452
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) .. (422)
; OTHER INFORMATION: n = g, a, c or t
; US-10-057-475B-3452

Query Match      7.0%; Score 16; DB 12; Length 422;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      154 ATTATCTCTCAGGAG 169
Db      291 ATTATCTCTCAGGAG 276

RESULT 41
US-10-154-884B-3452/C
; Sequence 3452, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
```



```
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3452
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(422)
; OTHER INFORMATION: n = g, a, c or t
US-10-154-884B-3452

Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ATTATCTCAGCGAG 169
Db 291 ATTATCTCAGCGAG 276

RESULT 42
US-10-040-862-3452/c
; Sequence 3452, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OR INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040, 862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3452
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (151)
; OTHER INFORMATION: n=A,T,C or G
```

```
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (165)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (340)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (347)
; OTHER INFORMATION: n=A,T,C or G
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (356)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-3452

Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ATTATCTCAGCGAG 169
Db 291 ATTATCTCAGCGAG 276

RESULT 43
US-09-796-692-2901
; Sequence 2901, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OR INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2901
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (151)
; OTHER INFORMATION: n=A,T,C or G

Query Match
7.0%; Score 16; DB 10; Length 427;
```

Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TTTATTAAGCACTCAA 24
|||||
Db 161 TTTATTAAGCACTCAA 176

RESULT 44

US-10-057-475B-2901
; Sequence 2901, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane A.
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aljun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2901
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-2901

Query Match 7.0%; Score 16; DB 12; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TTTATTAAGCACTCAA 24
|||||
Db 161 TTTATTAAGCACTCAA 176

RESULT 45

US-10-154-884B-2901
; Sequence 2901, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane

; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2901
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-2901

Query Match 7.0%; Score 16; DB 12; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TTTATTAAGCACTCAA 24
|||||
Db 161 TTTATTAAGCACTCAA 176

Search completed: February 4, 2004, 16:36:28
Job time : 1691 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 4, 2004, 14:31:59 ; Search time 2215 Seconds
(without alignments)
2501.770 Million cell updates/sec

Title: US-09-675-650-1_COPY_27_254

Perfect score: 228
Sequence: 1 gtaagtcgttataaagcac.....ttgtccctccacttcacag 228

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estbda:*
2: em_estbda:*
3: em_estbda:*
4: em_estbda:*
5: em_estbda:*
6: em_estbda:*
7: em_estbda:*
8: em_estbda:*
9: gb_est1:*
10: gb_est1:*
11: gb_est1:*
12: gb_est1:*
13: gb_est1:*
14: gb_est1:*
15: em_estfun:*
16: em_estfun:*
17: em_gss_hum:*
18: em_gss_hum:*
19: em_gss_hum:*
20: em_gss_hum:*
21: em_gss_hum:*
22: em_gss_hum:*
23: em_gss_hum:*
24: em_gss_hum:*
25: em_gss_hum:*
26: em_gss_hum:*
27: em_gss_hum:*
28: gb_gss1:*
29: gb_gss1:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	8.3	211	AV341547	AV341547 AV341547
2	19	8.3	389	BU465599	BU465599 603773567
3	19	8.3	394	AA097151	AA097151 0299F Pyr
4	19	8.3	487	AQ207356	AQ207356 HS_3240_B

5	19	8.3	619	BU441045	BU441045 604142968
6	19	8.3	672	BU368119	BU368119 603784821
7	19	8.3	678	BU165567	BU165567 Dario rer
8	19	8.3	679	BX263205	BX263205 BX263205
9	19	8.3	683	BB621652	BB621652 BB621652
10	19	8.3	688	BU255698	BU255698 603748787
11	19	8.3	716	BU357330	BU357330 603473484
12	19	8.3	757	BU470147	BU470147 603363138
13	19	8.3	770	BU378628	BU378628 603808560
14	19	8.3	787	BU268139	BU268139 603505112
15	19	8.3	792	BU446400	BU446400 603213673
16	19	8.3	810	BU114014	BU114014 603541537
17	19	8.3	833	BZ253924	BZ253924 CH230-499
18	19	8.3	850	BU122081	BU122081 603146589
19	19	8.3	873	BU121203	BU121203 603003002
20	19	8.3	873	BU387613	BU387613 603861170
21	19	8.3	873	BZ255847	BZ255847 CH230-382
22	19	8.3	943	BF980523	BF980523 602304270
23	19	8.3	951	CNS02BYW	AL130373 Telradon
24	19	8.3	1276	CC317416	CC317416 TAM32-27L
25	18	7.9	238	BB562950	BB562950 BB562950
26	18	7.9	266	BB564795	BB564795 BB564795
27	18	7.9	456	BE230315	BE230315 99AS667 R
28	18	7.9	473	AL386985	AL386985 MCB39B07
29	18	7.9	477	BE230459	BE230459 99AS575 R
30	18	7.9	514	BE942826	BE942826 EST424405
31	18	7.9	518	CC407239	CC407239 PUEGJ78TD
32	18	7.9	527	AZ700056	AZ700056 RPTC-23-2
33	18	7.9	548	CA390726	CA390726 G8112H07
34	18	7.9	550	AZ399438	AZ399438 IM0165U13
35	18	7.9	580	AM687310	AM687310 NF008C04R
36	18	7.9	589	AG242176	AG242176 Lotus jap
37	18	7.9	601	CB934904	CB934904 r98e06.Y
38	18	7.9	634	CA922246	CA922246 EST639964
39	18	7.9	655	AF074623	AF074623 AF074623
40	18	7.9	655	AI110716	AI110716 HA0156 Hu
41	18	7.9	659	AG242178	AG242178 Lotus jap
42	18	7.9	665	BZ526580	BZ526580 CGAH24ATC
43	18	7.9	674	BB886033	BB886033 LB00548A
44	18	7.9	704	CA922247	CA922247 EST639965
45	18	7.9	807	BG470649	BG470649 602511574

ALIGNMENTS

RESULT 1
AV341547/c
LOCUS
DEFINITION
AV341547 RIKEN full-length enriched, adult male olfactory bulb Mus
musculus cDNA clone 643052K10 3', mRNA sequence.

ACCESSION
AV341547
VERSION
AV341547.1 GI:6381599

KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211)

Komuro, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Iehi, Y., Iehikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai
C., Kawai, T., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y.,
Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,
Yoshiki, A., Yoshino, M., Yamamoto, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Komuro, H., et al. 1999)
Unpublished
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsunura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)
Itoch, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
19-44 (1999)
Please visit our web site (<http://genome.riken.go.jp/>) for
further details.

FEATURES

SOURCE

```

BASE COUNT
ORIGIN

62 a      28 c      33 g      88 t
BamH1"

/organism="Mus musculus"
/mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6450522K10"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male
olfactory bulb"
/notes="Site_1: SalI; Site_2: BamH1; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGGAGAGGATCCCAAGAGCTCTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGGAGATTCGAGTTATTAATTAATCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamH1 sites.
Vector: a modified pluscript KS(+) after bulk excision
from Lambda Ptc I. Cloning sites, 5' end: SalI; 3' end:
BamH1"

```

Query Match	8.3%	Score 19	DB 9	Length 211
Bates Local Similarity	100.0%	Pred. No.	33	
Matches	19	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
QY	208	ATTGTCTCTCCACTTTCAC	226	
Db	19	ATTGTCTCTCCACTTTCAC	1	
RESULT 2				
LOCUS	BU465589	389 bp	mRNA	EST 29-NOV-2002
DEFINITION	603773567P1 CSEQRBN19 Gallus gallus CDNA clone ChbST710f17 5', mRNA			
ACCSSION	sequence.			
VERSION	BU465589			
KEYWORDS	BU465589.1	GI:25955073		
	EST.			

SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	MEDLINE	COMMENT
	<i>Gallus gallus</i> (chicken)					
	<i>Gallus gallus</i>					
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Eucleostomi; Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.					
		1 (bases 1 to 389)				
		Boardman, P.E., Sant-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.				
		A Comprehensive Collection of Chicken CDNs				
		Curr. Biol. 12 (22), 1965-1969 (2002)				
		22335534				
		12445392				
		Contact: Simon Hubbard				
		Department of Biomolecular Sciences				
		University of Manchester Institute of Science and Technology (UMIST)				
)				
		PO Box 88, Manchester, M60 1OD, UK				
		Tel: 01612008930				
		Fax: 01612360409				
		Email: Simon.Hubbard@umist.ac.uk.				
FEATURES		Location/Qualifiers				
SOURCE		1..389				

```

/strain="layer"
/db xref="taxon:9031"
/clone="CHEST710E17"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1fb="CSEORBN19"
/notes="Organ: ovary; Vector: pbluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer using methyalted C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntend, ligated to NotI adapters, digested with EcoRI , size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pbluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."
BASE COUNT      99 a      85 c      109 g      95 t      1 others
ORIGIN
Query Match      8.3%; Score 19; DB 13; Length 389;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 3	
AA097151	
LOCUS	
DEFINITION	0299F <i>Pyrococcus furiosus</i> 1-2AP II library, F Robb <i>Pyrococcus</i>
ACCESSION	<i>furiosus</i> CDNA clone 0299, mRNA sequence.
VERSION	AA097151
KEYWORDS	AA097151.1 GI:1642747
SOURCE	EST.
ORGANISM	<i>Pyrococcus furiosus</i>
	<i>Pyrococcus furiosus</i>
	Archaea; Euryarchaeota; Thermococci; Thermococcales;
	Thermococcaceae; <i>Pyrococcus</i> .
REFERENCE	1 (bases 1 to 394)
AUTHORS	Bogger, K.M., Brummett, S.R., Boggett, A., Davis, M.C., Hujer, K.M., Domke,
	S.T., Szasz, J., Kavel, J., DiRuggiero, J., Fuller, C., Chase, J.W. and
	Robb, F.T.

TITLE A Survey of the Genome of the Hyperthermophilic Archaeon, *Pyrococcus furiosus*

JOURNAL Unpublished (1996)

COMMENT Contact: Brummet SR, Robb F
Brummet, Sequencing R&D Inc.
Amer sham Life Science, Inc.
Amer sham Life Science, 26101 Miles Rd, Cleveland, OH 44128
Tel: 2164649277; 4102348870
Fax: 2163600975; 4102348896
Email: cn288@cleveland.freenet.edu
Seq primer: T7.

FEATURES
source Location/Qualifiers
1. .394
/organism="Pyrococcus furiosus"
/mol_type="rRNA"
/strain="DSM 3638"
/db_xref="taxon:2261"
/clone="0299"
/lab_host="E. coli"
/note="Vector: 1-ZAP II; Site 1: Eco RI; Site 2: Eco RI; Genomic DNA was purified from P. furiosus cells and partially digested with Dra I, Eco RV, Hinc II, Pvu II, size selected, ligated to Eco RI linkers then cloned into the Eco RI site of 1-ZAP II, plasmid excision vector. Excision was performed in batch and individual clones retrieved by plating."

BASE COUNT 138 a 50 c 93 g 107 t 6 others

ORIGIN

Query Match 8.3%; Score 19; DB 9; Length 394;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCTTATTAAGCACTCAAT 25
|||||
89 GCTTATTAAGCACTCAAT 107

Db

RESULT 4 487 bp DNA linear GSS 18-SEP-1998
AQ207356 HS 3240 B1 G02 T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3240 Col=3 Row=N, genomic survey sequence.

LOCUS

DEFINITION

ACCESSION AQ207356.1 GI:3618561

VERSION

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 487)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3240 row: N column: 3
Class: BAC ends
High quality sequence stop: 487.
Location/Qualifiers

source 1. .487
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3240 Col=3 Row=N"
/sex="male"
/note="Vector: pBlueScript II KS(+); Site 1: EcoRI; Site 2: NotI. This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dt) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBlueScript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that

BASE COUNT 130 a 95 c 153 g 106 t 3 others

ORIGIN

Query Match 8.3%; Score 19; DB 28; Length 487;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 ATATTGTCCTCCACTTTC 224
|||||
394 ATATTGTCCTCCACTTTC 376

Db

RESULT 5 619 bp mRNA linear EST 29-NOV-2002
BU441045 604142968F1 CSBQRBN11 Gallus gallus cDNA clone CHEST978h18 5', mRNA sequence.

LOCUS

DEFINITION

ACCESSION BU441045

VERSION

KEYWORDS BU441045.1 GI:25930356

SOURCE EST.

ORGANISM Gallus gallus (chicken)

REFERENCE 1 (bases 1 to 619)
Bordman,P.B., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Pong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken CDNAS
Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .619
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST978h18"
/sex="Male and female"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_11b="CSBQRBN11"
/note="Vector: pBlueScript II KS(+); Site 1: EcoRI; Site 2: NotI. This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dt) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBlueScript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that

BASE COUNT 149 a 135 c 170 g 165 t
ORIGIN

a significantly longer reannealing hybridization was used."

Query Match 8.3%; Score 19; DB 13; Length 619;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 159 TCCTCAGGAGTCTGGATT 177
DB 41 TCCTCAGGAGTCTGGATT 59

RESULT 6
LOCUS BU368119 672 bp mRNA linear EST 28-NOV-2002
DEFINITION 603784821F1 CSEQCHN72 Gallus gallus CDNA clone CHEST73915 5', mRNA
SEQUENCE.
BU368119
ACCESSION
VERSION BU368119.1 GI:25876120
KEYWORDS
SOURCE EST.
ORGANISM Gallus gallus (chicken)

Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 672)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
JOURNAL MEDLINE
PUBMED 12445392

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers

1. .672
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST73915"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="CSEQCHN72"
/note="Organ: brain; Vector: Bluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dt) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

BASE COUNT
ORIGIN

Query Match 8.3%; Score 19; DB 13; Length 672;
Best Local Similarity 100.0%; Pred. No. 44;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 159 TCCTCAGGAGTCTGGATT 177
DB 54 TCCTCAGGAGTCTGGATT 72

RESULT 7
LOCUS BX165567 678 bp DNA linear GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEX-124L1, genomic survey sequence.
BX165567
ACCESSION
VERSION BX165567.1 GI:27997108
KEYWORDS
SOURCE GSS.
ORGANISM Danio rerio (zebrafish)

Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 678)
Humphray, S.J., Huckle, E. and Durham, J.L.
Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Unpublished
This sequence was generated from the T7 end of BAC 124L1. 124L1 is
part of the Daniokey BAC library created by R. Plaister and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/projects/D_rerio/
Location/Qualifiers

FEATURES

source

1. .678
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-124L1"
/tissue_type="testis"
/note="vector pindigobAC-536"
BASE COUNT 172 a 120 c 128 g 258 t
ORIGIN

Query Match 8.3%; Score 19; DB 29; Length 678;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 119 TTCTTGCCTCGATTGTC 137
DB 643 TTCTTGCCTCGATTGTC 661

RESULT 8
LOCUS BX263205 679 bp mRNA linear EST 27-FEB-2003
DEFINITION BX263205 AGENAE Gallus gallus multi-tissues normalized and
once-subtracted cDNA library (gcal) Gallus gallus cDNA clone
gcal0008c.n.07 5prim, mRNA sequence.
BX263205
ACCESSION
VERSION BX263205.1 GI:28585803
KEYWORDS
SOURCE EST.
ORGANISM Gallus gallus (chicken)

Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 679)
Herrault, F., Le Meuth-Metzinger, V., Desert, C., Retout, E., Plum, F.,
Klopp, C. and Douaire, M.
Construction and primary characterization of chicken normalized
multi-tissue cDNA libraries
Unpublished
Contact: Douaire M
INRA, UMR INRA-ENSAR Genetique Animale
65, rue de Saint-Brieuc, RENNES cedex, 35042, FRANCE
Tel: +33 (0) 2.23.48.54.63

JOURNAL
COMMENT

KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 698)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
2235534
JOURNAL MEDLINE
PUBMED
12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 698
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST657n21"
/dev_stage="36"
/lab_host="DH10B"
/note="Organ: limbs; Vector: Bluescript II KS(+); Site_1:
Ecotri; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

BASE COUNT 167 a 151 c 198 g 182 t
ORIGIN

Query Match 8.3%; Score 19; DB 13; Length 698;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGGAGTCTGGATT 177
|||||
Db 86 TCCTCAGGAGTCTGGATT 104

RESULT 11
BU357330 716 bp mRNA linear EST 28-NOV-2002
LOCUS 603473484F1 CSEQCHN70 Gallus gallus cDNA clone CHEST353g22 5', mRNA
DEFINITION
sequence.
ACCESSION BU357330
VERSION BU357330.1 GI:25865331
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 716)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 2235534
PUBMED 12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 716
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST353g22"
/dev_stage="36"
/lab_host="DH10B"
/note="Organ: hearts; Vector: pBluescript II KS(+);
Site_1: EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

BASE COUNT 171 a 159 c 203 g 183 t
ORIGIN

Query Match 8.3%; Score 19; DB 13; Length 716;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGGAGTCTGGATT 177
|||||
Db 105 TCCTCAGGAGTCTGGATT 123

RESULT 12
BU470147 757 bp mRNA linear EST 30-NOV-2002
LOCUS 603363138F1 CSEQRBN21 Gallus gallus cDNA clone CHEST258k9 5', mRNA
DEFINITION
sequence.
ACCESSION BU470147
VERSION BU470147.1 GI:25963724
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 757)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
2235534
JOURNAL MEDLINE
PUBMED 12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK.

FEATURES
SOURCE 1. .757

```

/organism="Gallus gallus"
/mol_type="mRNA"
/strain="layer"
/db_xref="taxon:9031"
/clone="CHEST258K9"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="CSEQRBN21"
/note="Organ: ovary; Vector: bluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dt) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntend, ligated to NotI adapters, digested with EcoRI and size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

```

	Query March	8.3%;	Score 19;	DB 13;	Length 757;	
	Best Local Similarity	100.0%;	Pred. No. 46;			
	Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	159 TCCTCAGCGAGTCTGGATT	177				
Dd	99 TCCTCACGAGTCTGGATT	117				
RESULT 13						
BU378628						
LOCUS	770 bp	mRNA	linear	EST 28-NOV-2002		
DEFINITION	603808560F1 CSEQCHN74 Gallus gallus CDNA clone ChEST791g16 5', mRNA					
SEQUENCE.						
ACCESSION	BU378628					
VERSION	BU378628..1 GI:25886629					
KEYWORDS	EST.					
SOURCE	Gallus gallus (chicken)					
ORGANISM	Gallus gallus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 770) Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickler,C., Brown,W.R.A., Wilson,S.A., and Hubbard,S.J. A Comprehensive Collection of Chicken cDNAs Curr. Biol. 12 (22), 1965-1969 (2002) 22335534 12445392 Contact: Simon Hubbard					
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PubMed						
COMMENT						

```

) University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
location/Qualifiers
1. 770
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
FEATURES
source

```

```

/db xref="taxon:9031"
/clone="CHEST791g16"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1ib="CSEOCCHN74"
/notes="Organ: kidney & adrenal; Vector: pBluescript II
KS(+), Site_1: EcoRI, Site_2: NotI; This normalized
library was constructed from 1 million independent clones.
cDNA synthesis was initiated using an oligo(dT) primer,
using methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT      186 a      171 c      206 g      207 t
ORIGIN
Query Match      8.3%; Score 19; DB 13; Length 770;
Best Local Similarity 100.0%; Prid. No. 46;
Matches 19; Conservative 0; Mismatches 0; Gaps 0;
159 TCCTCAGCGAGCTCGATT 177
|||||
|||
37 TCCTCAGCGAGCTCGATT 55

```

RESULT 14
 BU268139
 LOCUS
 DEFINITION
 BU268139
 60350511221 CSEQCHN52 Gallus gallus CDNA clone CHEST428H4 5', mRNA
 sequence.
 BU268139
 BU268139.1 GI:25539089
 EST.
 Gallus gallus (chicken)
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 787)
 Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 CONTACT: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
)
 PO Box 88, Manchester, M60 1OD, UK
 Tel: 01612008930
 Fax: 0161360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..787

```

/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST428h4"
/dev_stage="z3"
/lab_host="RDH10B"
/clone_1fb="CSECHN52"
/noise="Organ: limbs; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA

```

synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 188 a 177 c 211 g 211 t

Query Match 8.3%; Score 19; DB 13; Length 787;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGGAGTCTGGATT 177
|||||
54 TCCTCAGGAGTCTGGATT 72

RESULT 15 BU446400 792 bp mRNA linear EST 29-NOV-2002
LOCUS 60321367F1 CSEQRBN13 Gallus gallus cDNA clone CHEST196b19 5', mRNA
DEFINITION sequence.

ACCESSION BU446400
VERSION BU446400.1 GI:25935711
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 792)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392

COMMENT Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source Location/Qualifiers

1..792
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clone="CHEST196b19"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRBN13"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 188 a 186 c 229 g 189 t

QY 159 TCCTCAGGAGTCTGGATT 177
|||||
728 TCCTCAGGAGTCTGGATT 746

RESULT 16 BU314014 810 bp mRNA linear EST 28-NOV-2002
LOCUS 603541537F1 CSEQRHN61 Gallus gallus cDNA clone CHEST511j10 5', mRNA
DEFINITION sequence.

ACCESSION BU314014
VERSION BU314014.1 GI:25822015
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 810)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392

COMMENT Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source Location/Qualifiers

1..810
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 15T"
/db_xref="taxon:9031"
/clone="CHEST511j10"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRHN61"
/note="Organ: heart; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 188 a 192 c 236 g 194 t

Query Match 8.3%; Score 19; DB 13; Length 810;
Best Local Similarity 100.0%; Pred. No. 47;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGGAGTGTGATT 177
 |||||
 Db 613 TCCTCAGGAGTGTGATT 631

RESULT 17
 BZ253924 833 bp DNA linear GSS 15-OCT-2002
 LOCUS CH230-499B10.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
 DEFINITION CH230-499B10, genomic survey sequence.
 ACCESSION BZ253924
 VERSION BZ253924.1 GI:23957862
 KEYWORDS GSS.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 833)
 Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn
 'A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
 Jong, P. and Fraser, C.M.
 Rat BAC End Sequences from Library CHORI-230 MboI segment
 Unpublished
 Other GSSs: CH230-499B10.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (<http://www.chori.org/bacpac/orering/information.htm>). BAC end
 page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
 Plate: 499 row: B column: 10
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1..833
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SaNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-499B10"
 /sex="Female"
 /cell_type="Brain"
 /clone_id="CHORI-230 Segment 2"
 /note="Vector: pTARAC1.3; Site 1: MboI; Site 2: MboI;
 CHORI-230 Rat (BN/SaNHsd/MCW) BAC library produced by
 Pieter de Jong"

BASE COUNT 269 a 165 c 170 g 229 t

ORIGIN

Query Match 8.3%; Score 19; DB 29; Length 833;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GATGGCCTTAAGTTCCTCT 62
 |||||
 Db 174 GATGGCCTTAAGTTCCTCT 192

RESULT 18
 BUI22081 850 bp mRNA linear EST 25-NOV-2002
 LOCUS 603146589F1 CSECHL17 Gallus gallus cDNA clone CHEST146b24 5', mRNA
 DEFINITION sequence.

ACCESSION BUI22081
 VERSION BUI22081.1 GI:25332239
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 850)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335334
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..850
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHEST146b24"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_id="CSECHL17"
 /note="Organ: Kidney + adrenal; Vector: pBluescript II
 KS(+); Site 1: EcoRI; Site 2: NotI; Modification of
 pBluescript II KS(+) (Stratagene) vector to accommodate
 cDNA produced with the T-primed protocol (Construction of
 uni-directionally cloned cDNA libraries from messenger RNA
 for improved 3' end DNA sequencing by Glenn Fu, et al.
 U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
 NotI and EcoRI. Ligate in double stranded adaptor
 containing BspI and BamHI sites
 (5'ggcgcgtgcagcccgatccgaataaag)
 (5'aattcttttcgataccgggcgcgcgc)"

BASE COUNT 199 a 197 c 246 g 208 t

ORIGIN

Query Match 8.3%; Score 19; DB 13; Length 850;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGGAGTGTGATT 177
 |||||
 Db 539 TCCTCAGGAGTGTGATT 557

RESULT 19
 BUI21203 873 bp mRNA linear EST 25-NOV-2002
 LOCUS 603003002F1 CSECHL17 Gallus gallus cDNA clone CHEST13c2 5', mRNA
 DEFINITION sequence.
 ACCESSION BUI21203
 VERSION BUI21203.1 GI:25330906
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 873)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1OD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..873
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST13c2"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="CSEQCHL17"
/note="Organ: kidney + adrenal; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+) [Stratagene] vector to accommodate cDNA produced with the T-trimmed protocol (construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BspI and BamHI sites
[5'ggcggcgccagcccgatccgaataaaag] [5'aatctttcttcgagtcggcgctgcgc]"

BASE COUNT 209 a 192 c 245 g 227 t

ORIGIN

Query Match 8.3%; Score 19; DB 13; Length 873;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGCGAGCTCGATT 177
|||||
Db 225 TCCTCAGCGAGCTCGATT 243

RESULT 20
BU387613 873 bp mRNA linear EST 28-NOV-2002
LOCUS 603861170F1 CSEQCHN75 Gallus gallus cDNA clone CHEST872117 5', mRNA
DEFINITION
sequence.
ACCESSION BU387613
VERSION BU387613.1 GI:25895614
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 873)
Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1OD, UK
Tel: 01612008930
Fax: 01612360409

FEATURES
source
1..873
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, H1sex"
/db_xref="taxon:9031"
/clone="CHEST872117"
/dev_stage="36"
/lab_host="DH10B"
/clone_id="CSEQCHN75"
/note="Organ: trunks; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer. Using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adaptors, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 206 a 193 c 243 g 230 t 1 others

ORIGIN

Query Match 8.3%; Score 19; DB 13; Length 873;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TCCTCAGCGAGCTCGATT 177
|||||
Db 151 TCCTCAGCGAGCTCGATT 169

RESULT 21
B2255847 873 bp DNA linear GSS 15-OCT-2002
LOCUS CH230-382K20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION
B2255847
ACCESSION B2255847
VERSION B2255847.1 GI:23961629
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 873)
Zhao,S., Shetty,J., Shatsman,S., Tesgaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,B.F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 Mbci segment
Unpublished
Other GSSs: CH230-382K20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9112 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdjong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_end/rat/bac_end_intro.html
Seq primer: SP6
Plate: 382 row: K column: 20
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1. .873
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SENHed/MCW"
/db_xref="taxon:10116"
/clone="CH230-382K20"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARAC1.3; Site 1: Mbol; Site 2: Mbol;
CHORI-230 Rat (BN/SENHed/MCW) BAC library produced by
Pleter de Jong"
BASE COUNT 262 a 219 c 164 g 228 t
ORIGIN
Query Match 8.3%; Score 19; DB 29; Length 873;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 9 TTTATAAGCATCATTT 27
Db 761 TTTATAAGCATCATTT 779
RESULT 22
BF980523/c 943 bp mRNA linear EST 23-JAN-2001
LOCUS 602304270F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4395785 5',
DEFINITION mRNA sequence.
ACCESSION BF980523
VERSION BF980523.1 GI:12383335
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 943)
NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM0093 row: k column: 18
High quality sequence stop: 650.
Location/Qualifiers
1. .943
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4395785"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 88"
/note="Organ: small intestine; Vector: PCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 Kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 280 a 194 c 230 g 239 t
ORIGIN
Query Match 8.3%; Score 19; DB 10; Length 943;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 GGCTCACTGGATTATCC 161
Db 801 GGCTCACTGGATTATCC 783
RESULT 23
CNS02BW/c 951 bp DNA linear GSS 01-SEP-2000
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION 254C04 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL190373
VERSION AL190373.1 GI:7828477
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetradon.
REFERENCE 1
Roest Crolius, H., Jalllon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Broctier, P., Quetier, F.,
Saurin, W., and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
2
Roest Crolius, H., Jalllon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A., and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
10899143
3 (bases 1 to 951)
REFERENCE Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
Location/Qualifiers
1. .951
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="254C04"
/clone_1lb="g"
/note="Genoscope sequence ID : COAG254B02LP1-end : T7"
BASE COUNT 237 a 217 c 195 g 293 t 9 others
ORIGIN
Query Match 8.3%; Score 19; DB 29; Length 951;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 210 TTGTCTCCACCTTCACAG 228
Db 769 TTGTCTCCACCTTCACAG 751
RESULT 24
CC317416 1276 bp DNA linear GSS 14-MAY-2003
LOCUS TM32-27L22 Sp6.1 TAM32 Gallus gallus genomic clone TAM32-27L22,
DEFINITION genomic survey sequence.

ACCESSION CC317416
VERSION CC317416.1 GI:30711474
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 1276)
Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, M., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: SP6 ATTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 707.
Location/Qualifiers
1..1276
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-27L22"
/sex="female"
/cell_line="UCD001", inbred 256"
/clone_1lb="TAM32"
/note="Vector: pECBAC1; Site 1: EcoRI; Site 2: EcoRI;
TAM32 Female Chicken library - for library and clone
ordering information: <http://www.hbz.tamu.edu>"
BASE COUNT 379 a 324 c 197 g 376 t
ORIGIN
Query Match 8.3%; Score 19; DB 29; Length 1276;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 70 TCTATCTTCTCTACTACT 88
1249 TCTATCTTCTCTACTACT 1267
RESULT 25 238 bp mRNA linear EST 29-NOV-2000
BB562950 RIKEN full-length enriched, adult male heart Mus musculus
LOCUS BB562950/c
DEFINITION cDNA clone 1010001C13 5', mRNA sequence.
ACCESSION BB562950
VERSION BB562950.1 GI:11453842
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 238)
Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
Hasegaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Kojima, Y., Komio
Imotani, K., Ishii, Y., Itoh, M., Iwata, M., Kawai, J., Kojima, Y., Komio
H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K.,
Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A.,
Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
T., Toyota, T., Watanishi, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshitani, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Alizawa, K. et al. 2000)
TITLE Unpublished
JOURNAL
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermoregulation and thermoregulation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Iwata, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.
Location/Qualifiers
1..238
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="1010001C13"
/sex="male"
/tissue_type="heart"
/dev_stage="adult"
/lab_host="BNN132"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGAGAGAGAGCGCGCGCACTGCACTTTTCTTTTCTTTTCTTTT 3'. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence 15'
GAGAGAGAGAGAGCGCGCGCACTGCACTTTTCTTTTCTTTTCTTTT 3'.
cDNA was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC 1."
BASE COUNT 69 a 73 c 50 g 46 t
ORIGIN
Query Match 7.9%; Score 18; DB 10; Length 238;
Best Local Similarity 100.0%; Pred. No. 12e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 36 AAATTTTGATGCGCTTA 53
139 AAATTTTGATGCGCTTA 122
RESULT 26 286 bp mRNA linear EST 29-NOV-2000
BB564795 RIKEN full-length enriched, adult male stomach Mus
LOCUS BB564795/c
DEFINITION cDNA clone 2200001P22 5', mRNA sequence.
ACCESSION BB564795
VERSION BB564795.1 GI:11455687
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 286)
TITLE Unpublished
JOURNAL
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic

AUTHORS
 Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodojama, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Komoto, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toyota, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Aizawa, K. et al. 2000)

TITLE
 Unpublished
JOURNAL
 Contact: Yoshinide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thermotabilization and thermoreactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
source
 1. 286
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="2200001P22"
 /sex="male"
 /tissue_type="stomach"
 /dev_stage="adult"
 /lab_host="SOLR"
 /lab_lib="RIKEN full-length enriched, adult male stomach"
 /note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTTATTAATTAATTCATCCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."
BASE COUNT
 85 a 74 c 64 g 63 t
ORIGIN

Query Match 7.9%; Score 18; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 AAATTTTGATGGCTTA 53
Db 136 AAATTTTGATGGCTTA 119

RESULT 27
 BE230315 456 bp mRNA linear EST 07-JUN-2000
LOCUS
 99AS667 Rice Seedling lambda ZAPII cDNA library Oryza sativa
DEFINITION
 (indica cultivar-group) cDNA clone 99AS667, mRNA sequence.
ACCESSION
 BE230315
VERSION
 BE230315.1 GI:8956640
KEYWORDS
 EST.
SOURCE
 Oryza sativa (indica cultivar-group)
ORGANISM
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoideae; Oryza.
 1 (bases 1 to 456)
 Lee, M.C., Shin, Y.C., Lee, T.H., Jeong, S.H., Kim, J.K., Eun, M.Y. and Nam, B.H.

TITLE
 Large-scale Sequencing Analysis of ESTs from Rice Seedling
JOURNAL
 Unpublished
COMMENT
 Contact: Eun M.Y.
 Department of Cytoogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.asri.re.kr.
FEATURES
source
 1. 456
 Location/Qualifiers
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Milyang23"
 /db_xref="taxon:39946"
 /clone="99AS667"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /clone_lib="Rice Seedling lambda ZAPII cDNA library"
 /note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site"

BASE COUNT
 130 a 115 c 103 g 108 t
ORIGIN

Query Match 7.9%; Score 18; DB 10; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 CTTCCGTCCTCCATATT 210
Db 186 CTTCCGTCCTCCATATT 203

RESULT 28
 AL386985 473 bp mRNA linear EST 03-AUG-2000
LOCUS
 MCB39B07F1 MCB3 Medicago truncatula cDNA clone MCB39B07 T3, mRNA sequence.
ACCESSION
 AL386985
VERSION
 AL386985.1 GI:9686736
KEYWORDS
 EST.
SOURCE
 Medicago truncatula (barrel medic)
ORGANISM
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 473)
 Journeet, E.P., Crespeau, H., van-Tuijnen, D., Gouzy, J., Jallion, O.,
 Mibel, A., Carreau, V., Chatagnier, O., Kahn, D., Gnanmazzi-Pearson
 V. and Gamas, P.

TITLE
 Medicago truncatula ESTs from endomycorrhizal roots
JOURNAL
 Unpublished
COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Jouner, Laboratoire de
 Biologie Moléculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 M-est@toulouse.inra.fr Website :
 http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES

source

1. .473

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="Umalong"

/db_xref="taxon:3880"

/clone="MCBC39B07"

/tissue_type="arbuscular mycorrhiza"

/dev_stage="harvested 3 weeks post inoculation with Glomus

intraradices"

/clone_lib="MCBC"

/note="Vector: pBluescript PSK, Site_1: EcoRI, Site_2:

XhoI; M. truncatula sterilised seeds were germinated for

72h at 25 C, before transplanting into a 1/3 Epiloses soil

: 2/3 calcined Terragreen mix in the presence of onion

root fragments colonized by the arbuscular mycorrhizal

fungus Glomus intraradices (Schenck & Smith, isolate DP8

). The plants were watered every day and twice a week with

a modified nutrient Long Ashton solution without phosphate

but with a high level of nitrate. After 3 weeks RNA was

extracted from whole root systems. cDNA was prepared from

polyA+ enriched RNA. The cDNA was directionally ligated

into Uni-zap XR vector from Stratagene and packaged using

Gigapack Gold packaging extracts. Plasmids containing cDNA

inserts were mass-excised from phage stocks using Exasit

helper phage and propagated in SOLR cells. Clone ordering

and sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France). Note : EST may be of

Fungal origin."

BASE COUNT

161 a 63 c 149 t 2 others

ORIGIN

Query Match

Best Local Similarity 7.9%; Score 18; DB 9; Length 473;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

32 TCAGAAATTTTGATGCG 49

182 TCAGAAATTTTGATGCG 299

Db

RESULT 29

BE230459

LOCUS

DEFINITION

99AS575 Rice Seedling Lambda ZAPII CDNA Library Oryza sativa

(indica cultivar-group) cDNA clone 99AS575, mRNA sequence.

ACCESSION

BE230459

VERSION

BE230459.1 GI:8956544

KEYWORDS

EST.

SOURCE

ORGANISM

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehretaceae; Oryzae; Oryza.

1 (bases 1 to 477)

Lee M.C., Shin Y.C., Lee T.H., Jeong S.H., Kim J.K., Eun M.Y. and

Nahm B.H.

Large-scale Sequencing Analysis of ESTs from Rice Seedling

Unpublished

CONTACT: Eun M.Y.

Department of Cytogenetics

National Inst. of Agri. Sci. and Tech, RDA

Suwon, Kyungido, Korea

Tel: 82 331 290 0301

Fax: 82 331 290 0307

Email: myeun@sun20.asi.re.kr.

Location/Qualifiers

source

1. .477

/organism="Oryza sativa (indica cultivar-group)"

/mol_type="mRNA"

/cultivar="Milyang23"

/db_xref="taxon:3946"

/clone="99AS575"

/dev_stage="5 days after pollination"

/lab_host="E. coli SOLR"

/clone_lib="Rice Seedling Lambda ZAPII CDNA Library"

/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:

XhoI; Directional cDNA library inserted into lambda ZAPII

vector at 5' end with EcoRI and 3' end with Xho I site"

BASE COUNT

131 a 107 c 106 g 133 t

ORIGIN

Query Match

Best Local Similarity 7.9%; Score 18; DB 10; Length 477;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

193 CCTCCGTCCTCCATATT 210

186 CCTCCGTCCTCCATATT 203

Db

RESULT 30

BE942826

LOCUS

DEFINITION

BE942826

EST422405 MGHG Medicago truncatula cDNA clone pMGHG-9011, mRNA

sequence.

ACCESSION

BE942826

VERSION

BE942826.1 GI:10520585

KEYWORDS

EST.

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;

Medicago.

1 (bases 1 to 514)

Cote F., Ojamen-Reuths T., Hahn M.G., Vandenbosch K., Hur J.,

Beremand P., Endre G., Town C.D., Bowman C.L., Craven M.B. and Cho

J., Fraser C.M.

ESTs from seedling roots of Medicago truncatula after treatment

with beta glucan elicitor preparation from Phytophthora sojae

Unpublished

CONTACT: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

University of Georgia name: G270678e TIGR sequence name: MTJBA54TK

More information is available at: http://chrysle.tamu.edu/medicago

Seq primer: SKmod (CTA GAA CTA GTG GAT CC).

Location/Qualifiers

1. 514

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="pMGHG-9011"

/tissue_type="Roots from four day old seedlings"

/dev_stage="2 days after treatment with beta glucan

elicitor preparation from Phytophthora sojae"

/lab_host="E. coli strain XL0R"

/clone_lib="MGHG"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Uni-zap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

FEATURES

source

1. 514

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="pMGHG-9011"

/tissue_type="Roots from four day old seedlings"

/dev_stage="2 days after treatment with beta glucan

elicitor preparation from Phytophthora sojae"

/lab_host="E. coli strain XL0R"

/clone_lib="MGHG"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Uni-zap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

REFERENCE

AUTHORS

CONTACT: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

University of Georgia name: G270678e TIGR sequence name: MTJBA54TK

More information is available at: http://chrysle.tamu.edu/medicago

Seq primer: SKmod (CTA GAA CTA GTG GAT CC).

Location/Qualifiers

1. 514

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="pMGHG-9011"

/tissue_type="Roots from four day old seedlings"

/dev_stage="2 days after treatment with beta glucan

elicitor preparation from Phytophthora sojae"

/lab_host="E. coli strain XL0R"

/clone_lib="MGHG"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Uni-zap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

FEATURES

source

1. 514

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="pMGHG-9011"

/tissue_type="Roots from four day old seedlings"

/dev_stage="2 days after treatment with beta glucan

elicitor preparation from Phytophthora sojae"

/lab_host="E. coli strain XL0R"

/clone_lib="MGHG"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the Uni-zap XR vector from

Stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

helper phage and propagated in XL0LR cells."

BASE COUNT 172 a 68 c 95 g 177 t 2 others

ORIGIN

Query Match 7.9%; Score 18; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCAGAAATTTTGATGSC 49
|||||
Db 226 TCAGAAATTTTGATGSC 243

RESULT 31
LOCUS CC407239 518 bp DNA linear GSS 19-MAY-2003
DEFINITION PUEGJ78TD_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMWBTA246M12,
genomic survey sequence.
ACCESSION CC407239
VERSION CC407239.1 GI:30887329
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 518)
WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennettzen,J.
Maize Genomics Consortium
Unpublished
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
1. 518
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMWBTA246M12"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

BASE COUNT 187 a 104 c 109 g 118 t

ORIGIN

Query Match 7.9%; Score 18; DB 29; Length 518;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 TTTTCTATTTCTGCGCTC 129
|||||
Db 184 TTTTCTATTTCTGCGCTC 167

RESULT 32
LOCUS AZ700056 527 bp DNA linear GSS 24-JAN-2001
DEFINITION RPCI-23-223H8_TV RPCI-23 Mus musculus genomic clone RPCI-23-223H8,
genomic survey sequence.
ACCESSION AZ700056
VERSION AZ700056.1 GI:12421104
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 527)

REFERENCE
AUTHORS Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Aktinret
,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other GSSs: RPCI-23-223H8_TV
Contact: Shaying Zhao
Department of Bukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 223 row: H column: 8
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 527
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-223H8"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney end/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 114 a 103 c 98 g 212 t

ORIGIN

Query Match 7.9%; Score 18; DB 28; Length 527;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 TTCTACTCAGAAATTTT 43
|||||
Db 220 TTCTACTCAGAAATTTT 237

RESULT 33
LOCUS CA390726 548 bp mRNA linear EST 06-NOV-2002
DEFINITION csl12h07.y1 Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unamplified): cs Homo sapiens cDNA clone csl12h07
5', mRNA Sequence.
ACCESSION CA390726
VERSION CA390726.1 GI:24722000
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the
NIH Bank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
JOURNAL MEDLINE 22103460
PUBMED 12107410

COMMENT Contact: Wietow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: gwaemehelix.nih.gov
Plate: 112 row: h column: 07
Seq primer: M13RP1 reverse primer (ABI)

FEATURES
source
Location/Qualifiers
1..548
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ca112h07"
/tissue_type="RPE/choroid"
/dev_stage="adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA
(un-normalized, unamplified): c8"
/note="Organ: Eye; Vector: PCWVSORT6; Two different donor
eyes (75-80 years old) yielded approximately 600 mg of
dissected RPE/choroid tissue. This in turn yielded 340 ug
of total RNA and 7 ug of mRNA. A directionally cloned cDNA
library in the PCWVSORT6 vector was constructed at Life
Technologies (Rockville, MD; now part of Invitrogen Corp),
essentially following the protocols of the SuperScript
Plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/> The library code
designation was c8. For this library, cDNA inserts were
cloned into the NotI/MulI sites of the vector. EST
analysis was performed on the unamplified library at the
NIH Intramural Sequencing Center (NISC)."

BASE COUNT 157 a 100 c 107 g 184 t

ORIGIN

Query Match 7.9%; Score 18; DB 14; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 TTCTACTCACTGTCCTC 94
Db 115 TTCTACTCACTGTCCTC 132

RESULT 34
AC2399438 550 bp DNA linear GSS 03-OCT-2000
LOCUS M0165J13F Mouse 10kb plasmid UGCM library Mus musculus genomic
DEFINITION clone UGCM0165J13 F, genomic survey sequence.
ACCESSION AC2399438
VERSION AC2399438.1 GI:10514512
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 550)
REFERENCE 1 (bases 1 to 550)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,B., Pedersen,T., Reilly,
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
Plate: 0165 row: J column: 13
Seq primer: CGTGTGTAACGACGGCCACT
Class: plasmid ends
High quality sequence stop: 550.

FEATURES
source
Location/Qualifiers
1..550
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCM0165J13"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UGCM library"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydridynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
ligated DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (GI4732114|BP|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 209 a 105 c 109 g 127 t

ORIGIN

Query Match 7.9%; Score 18; DB 28; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 TTCTACTCAGAAATTTT 43
Db 153 TTCTACTCAGAAATTTT 136

RESULT 35
AW687310 580 bp mRNA linear EST 20-DEC-2000
LOCUS NF008C04RT1F1033 Developing root Medicago truncatula cDNA clone
DEFINITION NF008C04RT 5', mRNA sequence.
ACCESSION AW687310
VERSION AW687310.2 GI:11930471
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
Medicago.
1 (bases 1 to 580)
REFERENCE 1 (bases 1 to 580)
AUTHORS Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,
G.D. and Patva,N.L.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula root library
JOURNAL Unpublished
COMMENT On Apr 14, 2000 this sequence version replaced gi:7562046.
Contact: Patva NL
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7317

RESULT 38
LOCUS CA922246/c 634 bp mRNA linear EST 27-DEC-2002
DEFINITION EST639964 MTUS Medicago truncatula cDNA clone MTUS-51C5, mRNA
ACCESSION CA922246
VERSION CA922246.1 GI:27409176
KEYWORDS EST
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE 1 (bases 1 to 634)
Vandenbosch, K., Endre, G., Silverstein, K., Town, C.D., Van Aken, S., Utechtack, T., Cheung, F., and Fraser, C.M.
The Medicago truncatula 6K unigene set: cDNA clones selected and re-arrayed from various libraries
Unpublished
Contact: Vandenbosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenbosch@umn.edu
TIGR sequence name: MTUB29TV
More information is available at: www.medicago.org
Seq primer: (gta ata cga ctc act ata 999 C).

FEATURES
source
Location/Qualifiers
1..634
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="MTUS-51C5"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_id="MTUS"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

BASE COUNT 216 a 135 c 86 g 197 t
ORIGIN

Query Match 7.9%; Score 18; DB 14; Length 634;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 TCAGAAATTTTGATGCG 49
Db 253 TCAGAAATTTTGATGCG 236

RESULT 39
LOCUS AF074623/c 655 bp mRNA linear EST 29-NOV-1999
DEFINITION AF074623 Human fetal liver cDNA library Homo sapiens cDNA clone Ha0156, mRNA sequence.
ACCESSION AF074623
VERSION AF074623.1 GI:6470442
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 655)
AUTHORS Yu, Y., Zhang, S., Ouyang, S., Liu, M., Li, W., Bai, L., Wu, J., Zhou, S. and He, F.
TITLE Expression profile analysis of a human fetal liver cDNA library through large-scale sequencing
JOURNAL Unpublished
COMMENT Contact: Yongtao Yu
Department of Hematology
Beijing Institute of Radiation Medicine
27 Taiping Road, Beijing 100850, P.R.China
Tel: 0086-10-68159479
Fax: 0086-10-68214653
Email: yyt48@yahoo.com.

FEATURES
source
Location/Qualifiers
1..655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="Ha0156"
/tissue_type="liver"
/dev_stage="fetal"
/lab_host="MC1061/p3"
/clone_id="Human fetal liver cDNA library"
/note="Vector: pCDNA1"

BASE COUNT 222 a 118 c 144 g 171 t
ORIGIN

Query Match 7.9%; Score 18; DB 9; Length 655;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 ATTTCTATTCTTGCGCT 128
Db 27 ATTTCTATTCTTGCGCT 10

RESULT 40
LOCUS A110716/c 655 bp mRNA linear EST 11-NOV-1999
DEFINITION HA0156 Human fetal liver cDNA library Homo sapiens cDNA, mRNA sequence.
ACCESSION A110716
VERSION A110716.1 GI:6359581
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 655)
Yu, Y., Zhang, C., Luo, L., Ouyang, S., Li, W., Wu, J., Zhou, S., Liu, M. and He, F.
TITLE Expression profile analysis of a human fetal liver cDNA library
JOURNAL Unpublished
COMMENT Contact: Yongtao Yu
Department of Hematology
Beijing Institute of Radiation Medicine
27 Taiping Road, Beijing 100850, P.R.China
Tel: 0086-10-68159479
Fax: 0086-10-68214653
Email: yyt48@yahoo.com.

FEATURES
source
Location/Qualifiers
1..655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="liver"
/dev_stage="fetal"
/lab_host="MC1061/p3"
/clone_id="Human fetal liver cDNA library"
/note="Vector: pCDNA1"

BASE COUNT 222 a 118 c 144 g 171 t
ORIGIN

Query Match 7.9%; Score 18; DB 9; Length 655;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 ATTTCTATTCTTGCT 128
 |||||
 Db 27 ATTTCTATTCTTGCT 10

RESULT 41
 AG242178 659 bp DNA linear GSS 12-DEC-2002
 LOCUS Lotus japonicus DNA, clone:Lj10d15_sfi, genomic survey sequence.
 AG242178
 AG242178.1 GI:26553015
 GSS.
 SOURCE Lotus japonicus
 ORGANISM Lotus japonicus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
 Lotus.

REFERENCE 1
 Sato, S., Nakamura, Y. and Tabata, S.
 TITLE Lotus japonicus TAC End sequences
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 659)
 AUTHORS Sato, S.
 TITLE Direct Submission
 JOURNAL Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute,
 The First Laboratory for Plant Gene Research; 2-6-7
 Kazusa-Kametari, Kisarazu, Chiba 292-0818, Japan
 (E-mail: sato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/,
 Tel: 81-438-52-3935 (ex. 2335), Fax: 81-438-52-3934)
 Location/Qualifiers
 1. 659
 /organism="Lotus japonicus"
 /mol_type="genomic DNA"
 /strain="Miyakojima MG-20"
 /db_xref="taxon:34305"
 /clone="Lj10d15_sfi"
 /clone_lib="genomic TAC library"
 /note="VECTOR: pYLTACT"

BASE COUNT 227 a 116 c 74 g 242 t
 ORIGIN

Query Match 7.9%; Score 18; DB 29; Length 659;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 CAATTCTACTCAGAAAT 39
 |||||
 Db 464 CAATTCTACTCAGAAAT 481

RESULT 42
 BZ526580/c 665 bp DNA linear GSS 16-DEC-2002
 LOCUS OGAHP24TC ZM2_0.7_1.5 KB Zea mays genomic clone ZMMBMA0063C24,
 genomic survey sequence.
 BZ526580
 BZ526580.1 GI:27066595
 GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 665)
 whiteJaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick
 A., Frazer, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek
 R.W., Nunberg, A., Robbins, D. and Lakey, N.
 Consortium for Maize Genomics

JOURNAL Unpublished
 COMMENT Contact: Cathy WhiteJaw
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whiteJaw@ligr.org
 Seq primer: TP
 Class: shared ends.

FEATURES
 source location/Qualifiers
 1. 665
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBMA0063C24"
 /clone_lib="ZM2_0.7_1.5 KB"
 /note="Vector: pBCSK-, Site 1: HindIII, 0.7-1.5 kb
 methylation filtered genomic DNA library"

BASE COUNT 252 a 116 c 108 g 189 t
 ORIGIN

Query Match 7.9%; Score 18; DB 29; Length 665;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 TAAAGCAGCTCAATTCCTA 30
 |||||
 Db 236 TAAAGCAGCTCAATTCCTA 219

RESULT 43
 BH886033/c 674 bp DNA linear GSS 07-AUG-2002
 LOCUS LB00548a.d.SP6.1 Leishmania major Friedlin BAC library Leishmania
 DEFINITION major genomic clone LB00548a, genomic survey sequence.
 BH886033
 BH886033.1 GI:22130428
 GSS.
 SOURCE Leishmania major
 ORGANISM Leishmania major
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE 1 (bases 1 to 674)
 MYler, P.J., Vogt, C., Munden, H., Robertson, L., Sisk, E., Fazelinia, G.,
 Aggarwal, G., Nelson, S., Seyler, A., Mortley, E., Stuart, K. and
 Ragland, M.
 TITLE Leishmania major Friedlin BAC End Sequences
 JOURNAL Unpublished
 COMMENT Other GSSs: LB00548a.d.T7.1
 Contact: Myler, PJ
 Seattle Biomedical Research Institute
 4 Nickerson Street, Seattle, WA 98109-1651, USA
 Tel: 206 284-8846
 Fax: 206 284-0313
 Email: mylerpj@bri.org
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source location/Qualifiers
 1. 674
 /organism="Leishmania major"
 /mol_type="genomic DNA"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="LB00548a"
 /lib_host="E. coli GeneHogs + Trifa"
 /note="Vector: pCG270; Site 1: HindIII; Genomic DNA from
 Leishmania major Friedlin in agarose blocks was partially
 digested with HindIII, size selected, and ligated with
 HindIII-digested pCG270 vector DNA. 1038 clones were
 picked and arrayed in 384- and 96-well plates. Library
 construction and arraying was carried out by ResGen

Corporation and clones and filters are available from
them"

BASE COUNT 110 a 241 c 211 g 112 t
ORIGIN

Query Match 7.9%; Score 18; DB 28; Length 674;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 CTGACTGGCTCACTTGA 154
DB 350 CTGACTGGCTCACTTGA 333

RESULT 44
CA922247 704 bp mRNA linear EST 09-MAY-2003
LOCUS EST63965 MTUS Medicago truncatula cDNA clone MTUS-51C6, mRNA
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CA922247
CA922247
CA922247.1 GI:27409177
EST.
Medicago truncatula (barrel medic)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 704)
Vandenbosch, K., Endre, G., Silverstein, K., Town, C.D., Van Aken, S.,
Utrecht, T., Cheung, F., and Fraser, C.M.
The Medicago truncatula 6K unigene set: cDNA clones selected and
re-arrayed from various libraries

JOURNAL
COMMENT
Unpublished
Contact: Vandenbosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenbosch.umn.edu
Alias Clone PMHAM-19C1
TIGR sequence name: MTUB30TV
More information is available at: www.medicago.org
Seq primer: (gta ata cga ctc act ata ggg c).
Location/Qualifiers

FEATURES
SOURCE
1. .704
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultiivar="A17"
/db_xref="taxon:3880"
/clone="MTUS-51C6"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"

/note="Vector: pluescript SK-, Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-aseII
helper phage and propagated in XLOLR cells."

BASE COUNT 235 a 151 c 100 g 218 t
ORIGIN

Query Match 7.9%; Score 18; DB 14; Length 704;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCAGAAATTTTGTATGC 49
DB 253 TCAGAAATTTTGTATGC 236

RESULT 45
BG470649/c 807 bp mRNA linear EST 21-MAR-2001
LOCUS 602511574F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4634083 5',
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 807)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LICM1389 row: d column: 20
High quality sequence stop: 680.
Location/Qualifiers

FEATURES
SOURCE
1. .807
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4634083"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_16"
/note="Organ: eye; Vector: POTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCAAGG(G). Library constructed by Ling Hong
in the laboratory of Gerald W. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 225 a 176 c 206 g 200 t
ORIGIN

Query Match 7.9%; Score 18; DB 10; Length 807;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 TGCTGACTGGCTCACTT 151
DB 684 TGCTGACTGGCTCACTT 667

Search completed: February 4, 2004, 16:06:50
Job time : 2219 secs